Research Report 1347

IMPROVING THE SELECTION, CLASSIFICATION, AND UTILIZATION OF ARMY ENLISTED PERSONNEL:

ANNUAL REPORT

AD-A141 807

Human Resources Research Organization,
American Institutes for Research,
Personnel Decisions Research Institute,
and
Army Research Institute

SELECTION AND CLASSIFICATION TECHNICAL AREA

U.S. Army
Research Institute for the Behavioral and Social Sciences

October 1983

Approved for public release, distribution unlimited.

U. S. ARMY RESEARCH INSTITUTE
FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency under the Jurisdiction of the Deputy Chief of Staff for Personnel

L. NEALE COSBY
Colonel, IN
Commander

EDGAR M. JOHNSON
Technical Director

Technical review by:

NEWELL K. EATON
JOYCE L. SHIELDS

Notices

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: Research Report 1347	2. GOVT ACCESSION NO.: AD-A147 807 / 3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): Improving the Selection, Classification and Utilization of Army Enlisted Personnel: Annual Report	5. TYPE OF REPORT & PERIOD COVERED: Annual Report 1 Oct 82 - 30 Sep 83
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): Human Resources Research Organization American Institutes for Research Personnel Decisions Research Institute Army Research Institute	8. CONTRACT OR GRANT NUMBER(s): MDA 903-82-C-0531
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Human Resources Research Organization 1100 So. Washington Street Alexandria, VA 22314	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS: 20263731A792
11. CONTROLLING OFFICE NAME AND ADDRESS: U.S. Army Research Institute for the Behavioral and Social Sciences 5001 Eisenhower Ave., Alexandria, VA 22333	12. REPORT DATE: October 1983
	13. NUMBER OF PAGES: 89
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report): Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

The Army Research Institute technical point of contact is Dr. Newell K. Eaton. His telephone number is (202) 274-8275.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Validation	Job Knowledge Tests	Predictors
Predictor Measures	Validity Generalization	Criterion Measures
Construct Validation	Performance Measures	Army-wide Measures
Longitudinal Data Base		

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This research report describes the research performed during the first year of a project to develop a complete personnel system for selecting and classifying all entry level enlisted personnel. In general, the first year's activities have been taken up by an intensive period of detailed planning, briefing advisory groups, preparing initial troop requests, and beginning comprenensive predictor and criterion development that will be the basis for later validation work. A detailed description of the first year's work is contained in the Annual Report Technical Appendix, ARI Research Note 83-37.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Research Report 1347

IMPROVING THE SELECTION, CLASSIFICATION, AND UTILIZATION OF ARMY ENLISTED PERSONNEL:

ANNUAL REPORT

Human Resources Research Organization,
American Institutes for Research,
Personnel Decisions Research Institute,
and
Army Research Institute

Submitted by:

Newell K. Eaton, Chief
SELECTION AND CLASSIFICATION
TECHNICAL AREA

Approved as technically adequate
and submitted for publication by:

Joyce L. Shields, Director
MANPOWER AND PERSONNEL
RESEARCH LABORATORY

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES
5001 Eisenhower Avenue, Alexandria, Virginia 22333

Office, Deputy Chief of Staff for Personnel
Department of the Army

October 1983

Army Project Number
2Q263731A792

Manpower and Personnel

FOREWORD

This document describes the research performed during the first year of a project on a path toward achieving the goals of the Army's current, large-scale manpower and personnel research effort for improving the selection, classification, and utilization of Army enlisted personnel. The thrust for the project came from the practical, professional, and legal need to validate the Armed Services Vocational Aptitude Battery (ASVAB--the current US military selection/classification test battery) and other selection variables as predictors of training and performance. The portion of the effort described herein is devoted to the development and validation of Army Selection and Classification Measures, and referred to as "Project A." This work is funded primarily by Army Project Number 2Q263731A792. Another part of the effort is the development of a prototype Computerized Personnel Allocation System, referred to as "Project B." Together, these Army Research Institute research efforts, with their in-house and contract components, comprise a landmark program to develop a state-of-the-art empirically validated personnel selection, classification, and allocation system.

Edgar M Johnson

EDGAR M. JOHNSON
Technical Director, ARI and
Chief Psychologist, U. S. Army

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont.)

TABLES:

REFERENCES TO PAPERS, REPORTS, AND ARTICLES ABSTRACTED IN TEXT

REFERENCES TO PAPERS, REPORTS, AND ARTICLES ABSTRACTED IN TEXT (Cont.)

I. INTRODUCTION

The purpose of this annual report is to document a variety of technical aspects of the plans and accomplishments of Project A: Improving the Selection, Classification, and Utilization of Army Enlisted Personnel. Project A, together with the related Enlisted Personnel Allocation System research effort (Project B), is designed to provide a significant increase in Army readiness. These unique, long term, large scale research programs will tie together selection, classification and job allocation of Army enlisted personnel so that personnel decisions are made to optimize soldier performance and utilization of soldier skills and abilities. The research will provide information and procedures required to meet the manpower challenge of the coming decade by assuring that the most qualified people are enlisted, allocated, and retained. The objectives of the research are to develop an integrated personnel management system based on: 1) current and new personnel and performance measures, 2) accurate empirical prediction of future performance, 3) selection/classification, and MOS allocation at enlistment and reenlistment to optimize individual and system performance, and 4) what-if gaming to illustrate the performance impact of possible personnel management decisions.

The thrust of the program came from the practical, professional and legal need to demonstrate the validity of the Armed Services Vocational Aptitude Battery (ASVAB--the current military selection/classification test battery) and other selection variables used as predictors of training and job performance. Research planners at the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) realized the sizable resource commitment required to show ASVAB validity. The resource commitment would be mostly for the development and application of training and job performance measures. It became apparent that with moderate additional effort the predictor space could be greatly enhanced with new tests, and an integrated personnel management system could be developed to more optimally use the predictor and performance information.

Project Background

In response to Army, Congressional, and professional requirements, ARI began in 1980 to develop a major personnel selection, classification, and allocation research program. The basic requirement was to demonstrate the validity of the ASVAB as a predictor of both training and on-the-job performance. In reviewing the design needed to meet that requirement, the concept of a larger project began to emerge. With only a moderate amount of additional resources, new predictors in the perceptual, psychomotor, interest, temperament, and biodata domains could be evaluated as well. And a longitudinal research data base could be developed, linking soldiers' performance on a variety of variables from enlistment, through training, first tour assignments, reenlistment decisions, and for some, to their second tour. Finally, those data could be the basis for a new way to allocate personnel, making near-real-time decisions on the best match between characteristics of an individual enlistee or reenlistee and the requirements of available Army military occupational specialties (MOS).

To address the selection and classification portion of the effort, solicitation MDA 903-81-12-R-0158 "Project A: Development and Validation of Army Selection and Classification Measures" was issued Oct. 21, 1981. This milestone document can be viewed as the "official" starting point of this landmark research program which has now completed its first year. The program was intended to bring together the best Army in-house and contract research scientists in a combined effort to meet the Army's requirements for improving their enlisted personnel selection and classification processes and programs. In the solicitation, Army research psychologists mapped out a comprehensive 7-year effort to provide the tools and information necessary for implementation of a state-of-the-art selection and classification system for all enlisted personnel in the U.S. Army.

Changes at ARI

While the contract SOW and RFP process was ongoing, substantial changes were being made within ARI to increase emphasis in the manpower and personnel area. The new manpower and personnel laboratory was created, and Dr. Joyce L. Shields was chosen as director. To accommodate the substantial in-house portion of Project A, the selection and classification technical area was established, with Dr. Newell K. Eaton as chief. A major recruitment effort brought together a staff of experienced research scientists to execute the in-house research and to monitor the contract effort.

Formation of the Consortium

In anticipation of the solicitation (RFP), the presidents of the Human Resources Research Organization (HumRRO), American Institutes for Research (AIR), and Personnel Decisions Research Institute (PDRI) formed a consortium to develop a research proposal to meet the requirements of the forthcoming "Development and Validation of Army Selection and Classification Measures" Request for Proposal (RFP). It was agreed that HumRRO, as prime contractor, would assume responsibilities for overall contract management, technical direction and planning, and for scientific quality assurance. In response to the RFP, the consortium's proposal was submitted in January 1982. The contract was awarded to the HumRRO-AIR-PDRI consortium on September 30, 1982. The contract covered a 7-year research program at an estimated overall total cost of $16,390,000.

Project Outline

The overall purpose of Project A: Improving the Selection, Classification, and Utilization of Army Enlisted Personnel is to enhance the Army's ability to accomplish its peacetime and mobilization missions through improved matching of individuals to military occupational specialties (MOS). Specifically, Project A is to:

(1) validate existing selection measures against both existing and project-developed criteria, the latter to include both Army-wide performance measures based on

newly developed rating scales and direct measures of MOS-specific task performance;

(2) develop and validate new and/or improved selection and classification measures;

(3) validate proximal criteria, such as performance in training, as predictors of later criteria, such as job performance ratings, so that more informed reassignment and promotion decisions can be made throughout the individual's tour;

(4) determine the relative utility to the Army of different performance levels across MOS; and

(5) estimate the relative effectiveness of alternative selection and classification procedures in terms of their validity and utility for making operational selection and classification decisions.

The project must not be viewed and is not being conducted as a set of separate tasks that make "inputs" to one another and that are to be "integrated" somehow. Such a view misses the essential unity of the effort; Project A is one project and is organized into five major tasks.

Task 1. Validation

Task 1 has two major components. The first component is to maintain the data base and provide the analytic procedures to determine the degree to which performance in Army jobs is predictable from some combination of new or existing measures. The second component is to conduct the appropriate analyses to determine whether the existing set of predictors, new predictors, or some combination of new and existing predictors has utility over and above the present system. These two components must be accomplished using state-of-the-art technology in personnel selection research and data analytic methods.

Task 2. Developing Predictors of Job Performance

To date, a large proportion of the efforts of the armed services in this area have been concentrated on improving the ASVAB, which is now a well-researched, valid measure of general cognitive abilities. However, many critical Army tasks appear to require psychomotor and perceptual skills for their successful performance. Further, neither biodata nor motivational variables are now comprehensively evaluated. It is perhaps in these four noncognitive domains that the greatest potential for adding valid independent dimensions to current classification instruments is to be found. The objectives of Task 2 are to develop a broad array of new and improved selection measures and to administer them to three major validation samples. A critical aspect of this task is the demonstration of the incremental validity added by new predictors.

Task 3. Measurement of School/Training Success

The objective of Task 3 is to derive school and training performance indexes that can be used: (1) as criteria against which to validate the initial predictors, and (2) as predictors of later job performance. Comprehensive job knowledge tests will be developed for the sample of MOS investigated and their content and construct validity will be determined.

Task 4. Assessment of Army-wide Performance

In contrast to performance measures which may be developed for a specific Army MOS, Task 4 will develop measures that can be used across all MOS (i.e., Army-wide). The intent is to develop measures of first- and second-tour job performance against which all Army enlisted personnel may be measured. A major objective for Task 4 is to develop a model of soldier effectiveness that specifies the major dimensions of an individual's contribution to the Army as an organization. Another important objective of Task 4 is to develop measures of utility. It is critical to define, in dollar terms, the benefits likely to accrue from what will probably be more costly selection/classification procedures.

Task 5. Develop MOS-Specific Performance Measures

The focus of Task 5 is the development of reliable and valid measures of specific job task performance for a selected set of MOS. This task may be thought of as consisting of three major components: job analysis, construction of job performance measures, and construct validation of the new measures. While only a subset of MOS will be analyzed during this project, the Army may in the future wish to develop job performance measures for a larger number of MOS. For this reason, the methods are intended to apply to all Army MOS.

The Consortium/ARI Team

The initial project organization is shown in Figure 1. The principal consortium task scientists are shown, with their respective organizations, in the lower row. The principal ARI scientists are shown in the upper row. In the project consortium and ARI scientists undertake both independent and joint research activities. ARI scientists also have the administrative role of contract oversight.

During the course of this first year, the consortium's organization structure has remained stable. However, a number of significant personnel changes did occur. In July 1983, Dr. Joe Olmstead, after having completed his supervision of the work entailed in achieving the project's "Research Plan" and "Master Plan," asked to be relieved of his responsibilities in order to pursue other interests. Dr. Robert Sadacca assumed responsibilities as Task 4 Leader.

Technical and management oversight is the responsibility of Dr. Newell K. Eaton, the contracting officer's technical representative (COR). On the

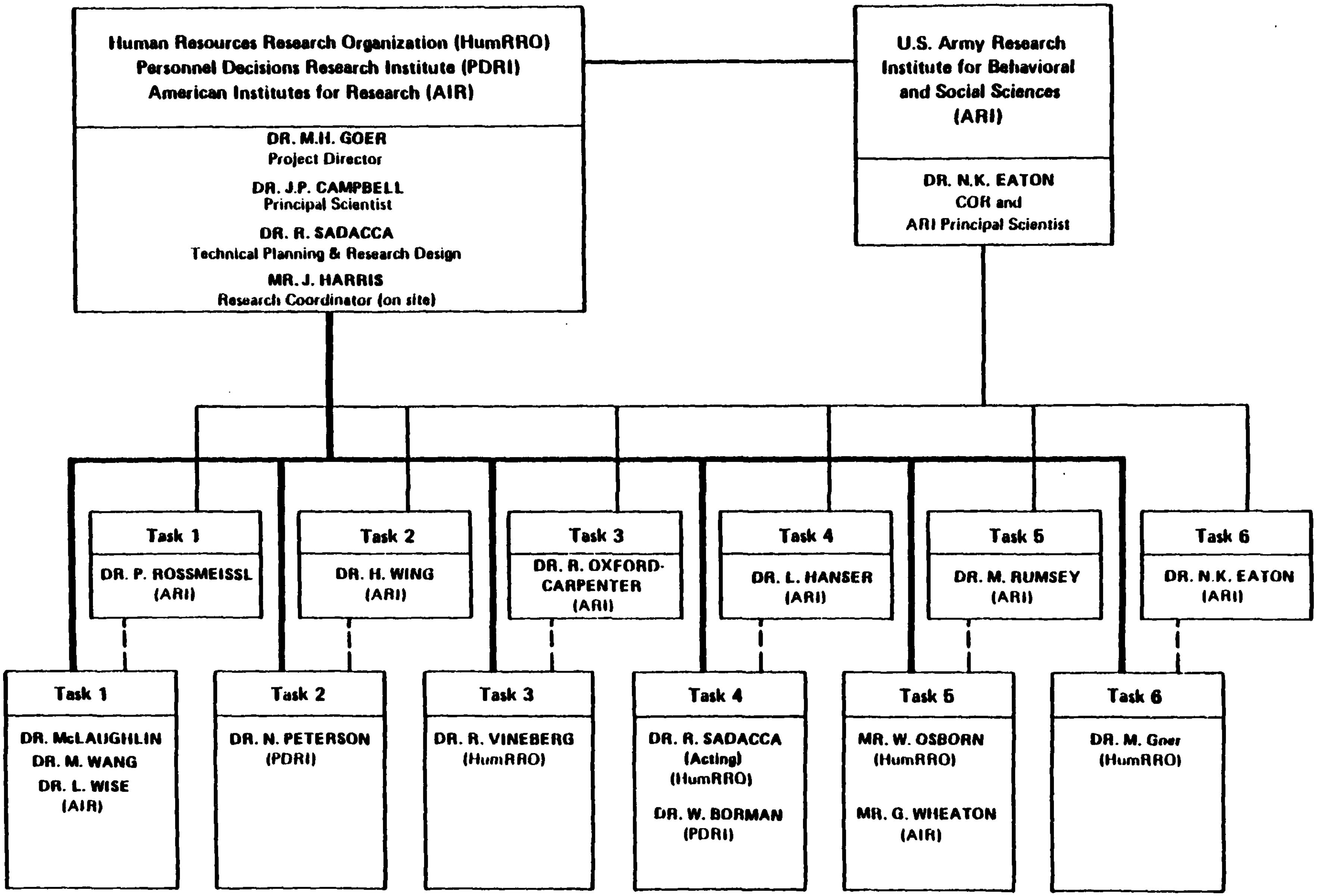


Figure 1. Project Organization

project he is the ARI principal scientist, and has responsibility for technical review and guidance for the consortium scientists and managers, as well as for the ARI research teams. Consortium management is the responsibility of Dr. Marvin H. Goer. He provides management functions to include planning, coordinating, and integrating. Dr. Goer is assisted in his role as Managing Project Director (MPD) by Dr. John P. Campbell, Dr. Robert Sadacca, and Mr. James Harris. Within the consortium, Dr. Campbell is the principal scientist responsible for overall scientific quality and for its state-of-the-art procedures. Dr. Sadacca is the assistant for technical planning and research design. In this role, he conceptualizes technical issues and integrates technical plans across tasks. Mr. Harris is the research coordinator on-site at ARI headquarters. As research coordinator, he conducts day-to-day liaison with the COR regarding Project A, Project B interactions, and related research.

The Advisory Group Structure

Because a program and project of this scale and importance would have to maintain close and active coordination with the other military departments, as well as with the Department of Defense, the project planners needed assurance that Project A was consistent with and complementary to other on-going research programs being conducted by the other armed services. The project also needed a mechanism for assuring that the research program met the highest standards for scientific quality and state-of-the-art technology in personnel selection and classification research. Finally, because it takes some time in a longitudinal research program to arrive at definitive answers to questions, a method was needed to receive feedback from senior officers on priorities and objectives, as well as to identify current problems where an appropriate research focus would bring operationally useful early results. An effective mechanism was essential because the research program involved large numbers of troops. Their commanders would require justification for use of those assets.

Figure 2 shows the structure and membership of the Governance Advisory Group (GAG) which comprises the Scientific Advisory Group (SAG), Interservice Advisory Group (ISAG), and Army Advisory Group (AAG) components. The SAG comprises nationally recognized authorities in psychometrics, experimental design, sampling theory, utility analysis, applied research in selection and classification, and in the conduct of psychological research in the Army environment. The ISAG comprises the Laboratory Directors for applied psychological research in the Army, Air Force, and the Navy, and the Director of Accession Policy from the DoD Office of Assistant Secretary of Defense for Manpower and Reserve Affairs. The AAG includes representatives from the Office of Deputy Chief of Staff for Personnel (DCSPER), Office of Deputy Chief of Staff for Operations (DCSOPS), Training and Doctrine Command (TRADOC), Forces Command, (FORSCOM), and U.S. Army Europe (USAREUR). These senior officers have a significant interest in the project planning and priorities. They also represent the elements which provide the necessary and substantial troop support.

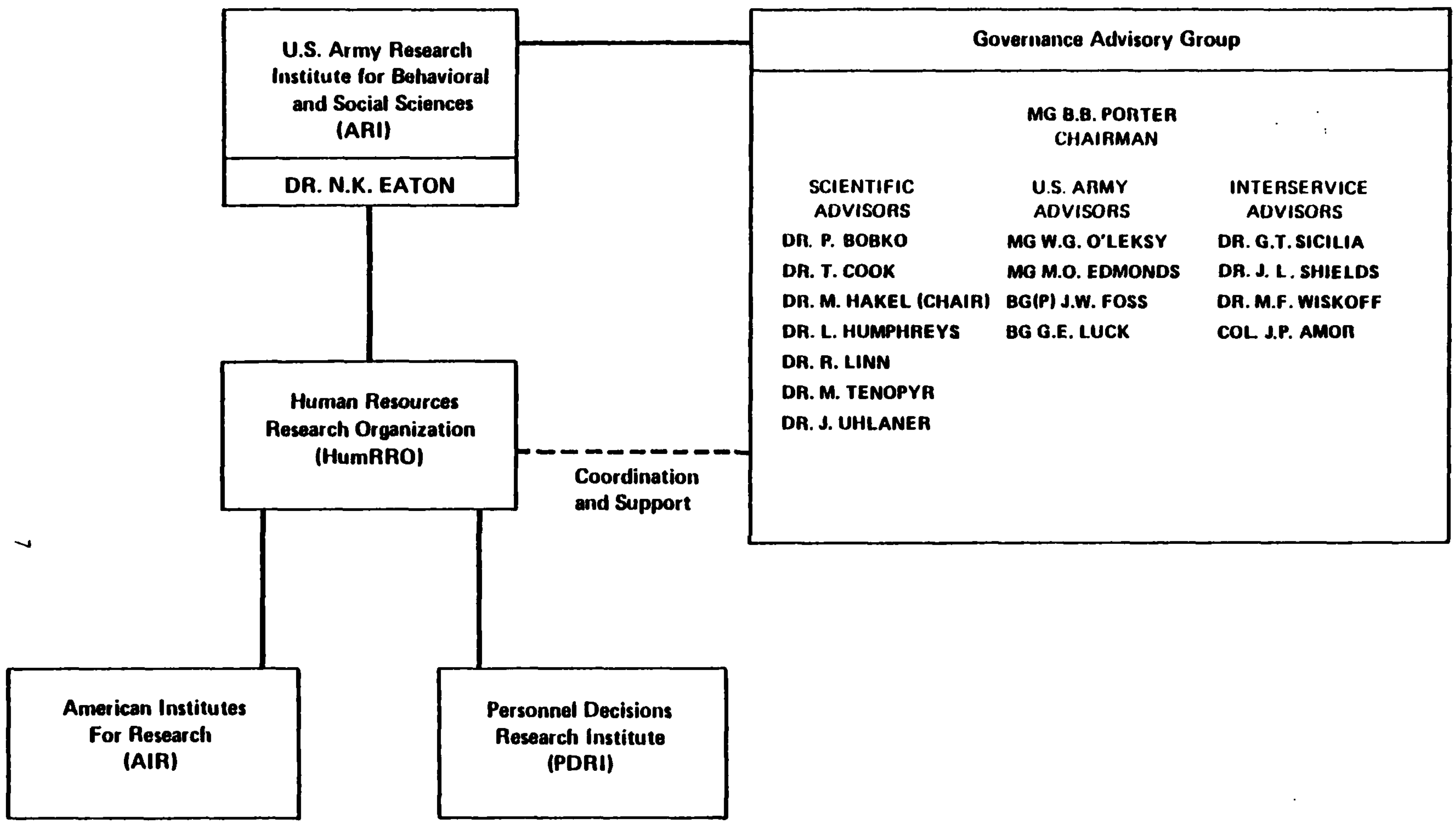


Figure 2. Governance Advisory Group

The Research Plan and Integrated Master Plan

The first six months of the project were spent in planning, documenting, reviewing, modifying, and redrafting of research plans, troop support, administrative support, and budgetary plans, as well as in execution of initial research efforts. Drafts of the plans were provided to the SAG and ISAG. Their comments, provided orally during meetings and subsequently written in response to draft documents, were addressed and their suggestions were incorporated in the research plan. The culminating review was conducted in April by the U.S. Army Advisory Group, with representatives from the Scientific and InterService Advisory Groups. In that meeting the entire research program, research design, sampling strategy, main cohort and focal MOS recommendation, and troop support implications were reviewed. Changes were incorporated to reduce and to distribute the troop support burden more equitably among the three participating commands (FORSCOM, TRADOC, USAREUR). The research program was endorsed by all three components of the GAG.

In May 1983, ARI issued Research Report 1332 "Improving the Selection, Classification, and Utilization of Army Enlisted Personnel - Project A: Research Plan." In June 1983, the "Project A: Integrated Master Plan" (HumRRO FR-PRD-83-8) was issued, providing detailed budget allocation, schedules, and product definitions.

In this first year a number of significant research activities were initiated and significant progress was made. The following sections of this report summarize some of the most important efforts through the period ending 30 September, 1983. Abstracts of associated research reports are included at the end of each section. A detailed description of the first year's work and the complete report for each abstract are contained in the Technical Appendix to this Annual Report, ARI Research Note 83-37 (Eaton & Goer, 1983).

General Outcomes

The Project A Research Plan speaks to the specific operational and scientific outcomes that will flow from the project. They are characterized by the following themes:

(1) Project A will generate a broader and more complete sample of the predictor space than has ever been used before in a selection investigation. The taxonomy of predictors that is established will stand as a reference point for many years to come.

(2) Project A will provide the most thorough attempt ever made to develop standardized tests of actual task performance in skilled jobs. The procedure used will stand as a model.

(3) Project A will be by far the most thorough test to date of whether success in training predicts success on the job.

(4) Project A will provide a state-of-the-art model to illustrate how construct validity can be used to study applied problems in selection/classification and performance assessment.

(5) Project A will be the first large selection and classification research effort to incorporate utility in the development of operational decision rules.

(6) Given the broad range of predictors, criteria, and jobs, Project A will be the most comprehensive evaluation ever conducted on questions of differential predictability across jobs, criterion measures, and predictor constructs.

We believe that Project A will make significant contributions to improve Army operational capability and to provide the most satisfactory careers for individual soldiers. Further, we expect that substantial scientific development will result from this effort. While it will be time consuming and expensive, in our judgment the benefits of this Project will be well worth the cost.

II. RESEARCH DESIGN AND SAMPLE SELECTION

Research Design

The overall design of Project A is described in detail in the Final Research Plan (June 1983). Briefly, the overall objectives are to develop and validate an experimental battery of new and improved selection measures against a comprehensive array of job performance and training criteria. The validation research must produce sample estimates of the parameters necessary to implement a computerized selection and classification system for all first-tour enlisted MOS. To do this a design was developed that uses two predictive and one concurrent validation on two major troop cohorts (83/84 accessions and 86/87 accessions), and one file data validation on the 81/82 cohort.

In addition to collecting data from new samples, the project is making use of existing file data that have been, or can be, accumulated for 1981 and 1982 accessions. The editing and merging of data from the accessions and EMF files for entry into the Longitudinal Research Data Base (LRDB) is now virturally complete and ready for analyses. A schematic of the data collection plan is shown in Figure 3.

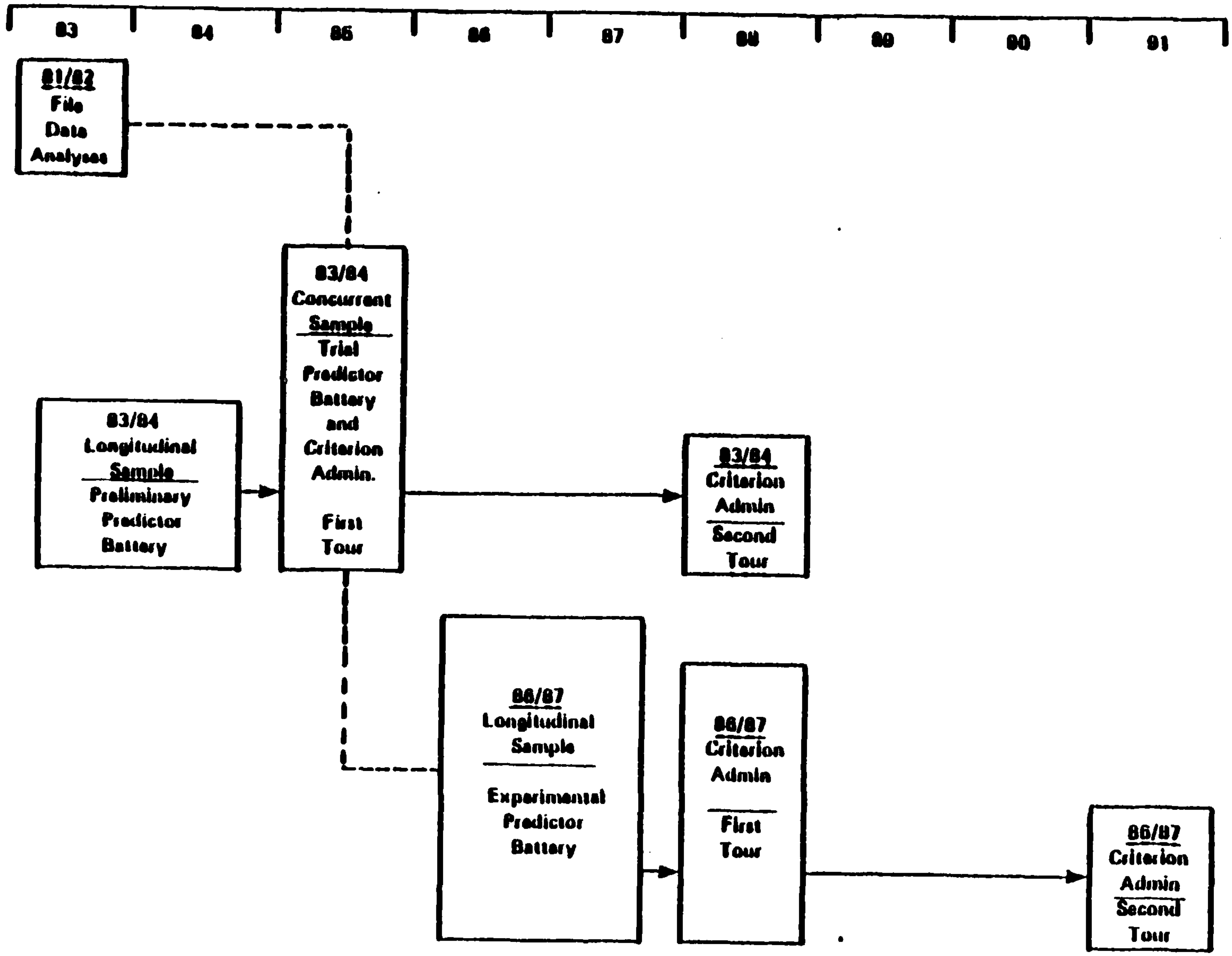
The logic of the design is straightforward. Existing file data on the 81/82 cohort provides an early opportunity to impact on the operational selection and classification system. Results will provide operational users with state-of-the-art solutions to selection and classification problems using the best available data. It will also point out analytic, operational, and policy difficulties to address and overscore in later efforts. The 83/84 cohort provides the first opportunity to obtain data using new predictor and performance measures. A "preliminary" battery of predominantly off-the-shelf tests provides new predictor data on soldiers in 4 MOS (05C, 19E/K, 63B, 71L). Development and refinement of these measures will provide a more tailored trial battery ready for application in 1985 concurrent with a variety of training, Army-wide, and MOS specific performance measures. The refinement of these measures will result in an experimental battery and a variety of criterion measures for application to a longitudinal sample of the FY86/87 cohort. In addition, second-tour Army-wide and MOS specific performance measures will be developed for and administered to both FY83/84 and FY86/87 cohort samples.

Sample Selection

Project A's large, complex requirements address simultaneously a wide range of interrelated research questions pertaining to an entire organizational personnel system. The overall objective in generating the samples has been to maximize the validity and reliability of the information to be gathered, while at the same time minimizing the time and costs involved. In part, costs are a function of the numbers of people in the sample. But, costs are also influenced by the relative difficulty involved in locating and assembling the people in a particular sample, by the degree to which the unit's operations are disrupted by the data collection, by the staff costs involved

Figure 3

The Overall Data Collection Plan



in collecting the data in a particular manner, and by other such considerations. However, cost considerations cannot be used to compromise the validity and statistical reliability of the data to the point where the necessary research and development questions cannot be answered with confidence. We have tried to balance these considerations in a feasible and appropriate way as the sampling plan was developed and implemented.

The sampling plan itself incorporated two principal considerations. First, a sample of MOS was selected from the universe of possible MOS. Then, the required sample sizes of enlisted personnel (EP) within each MOS were specified. The MOS are the primary sampling units. This is because Project A is developing a system for a population of jobs (MOS), but only a sample of MOS can be studied. Large and representative samples of enlisted personnel within each MOS are important because stable statistical results must be obtained for each MOS. There is a trade-off in the allocation of project resources between the number of MOS researched and the number of subjects tested within each MOS: the more MOS are investigated, the fewer subjects per MOS can be tested, and vice versa. Cost versus statistical reliability considerations dictated that 19 MOS could be studied. To samples from all 19, we will administer the new predictors (from Task 2) and collect the school and Army-wide performance data (of Tasks 3 and 4). To nine of these MOS, we will also administer the MOS-specific performance measures developed in Task 5. The nine MOS were chosen to provide maximum coverage of the total array of knowledge, ability, and skill requirements of Army jobs, given certain statistical constraints.

MOS Selection

The selection of the sample of 19 MOS proceeded through a series of stages. An initial sample of MOS was drawn by using the following considerations:

(1) High density MOS that would provide sufficient sample sizes for statistically reliable estimates of new predictor validity and differential validity across racial and gender groups.

(2) Representative coverage of the aptitude areas measured by the ASVAB area composites.

(3) High priority MOS (as rated by the Army in the event of a national emergency).

(4) Representation of the Army's designated Career Management Fields (CMF).

(5) Representation of the jobs most crucial to the Army's mission.

A further indirect indication of the mix of job skills represented in the sample is in the range of ASVAB composites and component subtest pertinent to each MOS. All subsets and all but one (EL) of the nine composites were represented in the 18 MOS initially selected. Consequently, a 19th MOS (27E) was chosen to represent the EL aptitude composite. The composition of

the sample was also examined from the standpoint of mission criticality by comparing it with a list of 42 MOS identified by the Army as high priority for mobilization training.[1]

This initial set of 19 MOS represent 19 of the Army's 30 CMF. Of the 11 CMF not represented, two are classified (CMF 96 and 98), two (CMF 33 and 74) have fewer than 500 FY81 accessions, and seven (CMR 23, 28, 29, 79, 81, 84, and 74) have fewer than 300 FY81 accessions. The initial set includes only 5 percent of Army jobs but 44 percent of the soldiers recruited in FY81. Similarly, of the 15 percent women in the 1981 cohort, 44 percent are represented in the sample; of the 27 percent Blacks, 44 percent are represented in the sample; and, of the 5 percent Hispanic, 43 percent are represented. Although female and minority representation is high absolutely, relatively it remains about the same as in the population. The sample is 15 percent female, 27 percent Black, and 5 percent Hispanic.

Nine of the 19 MOS were tentatively earmarked for the job specific performance measurement phase of the project. These were selected, as a subset, with the same general criteria used in identifying the parent list of 19. Since the larger list is composed of five combat and 14 noncombat MOS, it seemed reasonable that these categories were proportionally represented in the subset of nine.

Thus, the nine MOS designated for hands-on performance measurement development are:

(1) 11B - Infantryman
* (2) 13B - Cannon Crewman
(3) 19E/K - Tank Crewman
(4) 05C - Radio TT Operator
(5) 63B - Vehicle and Generator Mechanic
* (6) 64C - Motor Transport Operator
* (7) 71L - Administration Specialist
(8) 91B - Medical Care Specialist
* (9) 95B - Military Police.

An initial batch of four (see asterisks preceding) was selected and designated as Batch A; the other five as Batch B. Work has begun on Batch A first. Batch B will be taken up in turn.

On the basis of guidance from the Scientific Advisory Group, further refinements of the MOS sample were undertaken. These included a cluster analysis of expert ratings of MOS similarity and a review of the initial sample by the Governance Advisory Group.

Cluster Analysis

To obtain data for empirically clustering MOS on the basis of their task content similarity, a brief job description was generated for each of 111 MOS from the job activities described in AR 611-201. The sample of 111 MOS

[1]ODCSOPS (DAMO-ODM), DF, 2 Jul 82, Subject: IRR Training Priorities.

represents 47 percent of the population of 238 Skill Level 1, Active Army MOS with conventional ASVAB entrance requirements and includes the 84 largest MOS (300 or more new job incumbents yearly) plus an additional 27 selected randomly but proportionately by CMF. Each job description was limited to two sides of a 5x7 index card.

Members of the contractor research staff and ARI Army officers--approximately 25 in all--served as expert judges and were given the task of sorting the sample of 111 job descriptions into homogeneous categories based on perceived similarities and differences in job activities as described in AR 611-201. Data from the similarity scaling task were clustered and the initial results used to check the representativeness of the initial sample of 19 MOS. That is, did the initial sample of MOS include representatives from all the major clusters of MOS derived from the similarity scaling? On the basis of these results and guidance received from the Governance Advisory Group, two MOS that had been selected initially were replaced by 51B and 27E, which are in the same CMF and involve the same Aptitude Area Composites as the replaced MOS (62E and 31M).

The sample of MOS resulting from the above procedures is shown in Table 1.

Sampling Enlisted Personnel Within MOS

There are two major considerations relative to sampling individuals within MOS. One concerns the number of people per MOS and the other deals with the schedule or sampling plan for obtaining the data from the enlisted personnel serving as research subjects. The sampling plan, or design, is dictated by the research questions and the kind of information that is needed to answer them. The sample size within MOS is a function of the number of individuals needed for statistical reliability and the amount of sample attrition that must be allowed for to obtain such a sample size.

Summary

During the first year the focal MOS were selected, the sample sizes required from each were specified, and the troop support requests were prepared. In addition, the available computer file data on the 81/82 cohort were merged from the various sources, were thoroughly edited, and were made ready for analysis. Although the troop support requirements may seem large, they are made necessary by requirements of the selection and classification system to be developed. A series of smaller efforts over a longer period may indeed be more expensive in the end, and it would not produce the necessary data that Army management could use with confidence.

Abstract

As indicated in the Introduction, an abstract of a relevant research report follows.

Table 1

Project A MOS

MOS	Title	CMF	APT Comp	Priority MOS	Total	FY81 Accessions: Women	Blacks	Hispanic	Trainee Projections: FY83	FY84	Expected Number Graduates[1]
05C	Radio TT Operator	31	SC	No	3175	585	898	119	2004	2200	1645
63B	Vehicle & Generator Mech	63	MM	No	4663	386	1178	242	5304	4402	4280
64C	Motor Transport Operator	64	OF	Yes	6440	774	1279	141	3706	5000	4484
71L	Admin Specialist	71	CL	No	4484	2744	1967	215	6191	4592	3859
13B	Cannon Crewman	13	FA	Yes	5783	0	2053	367	6092	3553	3572
91B	Medical Care Specialist	91	ST	Yes	3074	924	876	224	3761	unav	3621
19E/K	Tank Crewman	19	CO	Yes	3233	0	604	168	3223	3261	2912
95B	Military Police	95	ST	Yes	6073	704	624	127	5720	5300	4373
11B	Infantryman	11	CO	Yes	7028	0	1128	367	12633	13710	11338
76Y	Unit Supply Specialist	76	CL	No	4565	1179	1998	283	6636	4091	3829
94B	Food Service Specialist	94	OF	No	3859	715	1416	125	5133	5157	4600
12B	Combat Engineer	12	CO	Yes	3707	0	716	147	844	2540	1845
16S	MANPADS Crewman	16	OF	Yes	691	0	206	27	797	1015	815
55B	Ammunition Specialist	55	GM	No	662	171	283	42	620	810	762
76W	Petroleum Supply Spec	92	CL	NO	849	259	559	43	1373	1350	1234
54E	Chemical Operations Spec	54	ST	Yes	557	89	185	41	1012	1247	1068
67N	Utility Helicopter Rpr	67	MM	No	1032	33	68	29	572	465	470
51B	Carpentry/Masonry Spc	51	GM	No	602	6	136	14	120	483	341
27E	Tow/Dragon Rpr	27	EL	No	333	40	76	17	312	308	258
	Total				59800	8609	16001	2758	66052	59511	55306

[1]Weighted average of Trainee Projections (3 months of FY83 and 9 months of FY84) adjusted for expected school attrition (actual FY81 rates).

GROUPING ARMY OCCUPATIONAL SPECIALTIES BY JUDGED SIMILARITY*

R. Rosse and W. Borman
(PDRI)
and
C. Campbell and W. Osborn
(HumRRO)

The purpose of this research was to group U.S. Army Military Occupational Specialties (MOS) into clusters according to rated similarities in MOS job content. These clusters were intended to guide a sampling of MOS selected to be representative of the entire Army MOS content domain. To accomplish this, 25 judges sorted the job descriptions of 111 entry-level MOS into categories based on perceived overall job similarity. An 111 x 111 pooled similarity matrix was formed, and correlations were computed between each pair of MOS on the basis of patterns of these similarities. A factor analysis of the correlation matrix yielded 15 orthogonal factors which were reinterpreted in a 23-factor oblique solution, with 1-14 MOS loading substantially on each factor. The paper discusses the resulting structure.

---

*Paper presented at the 25th Annual Conference of the Military Testing Association in Gulf Shores, Alabama, October 1983.

III. GENERATION OF THE LONGITUDINAL RESEARCH DATA BASE (LRDB) AND FY81/82 DATA FILE

Project A will generate a large amount of interrelated data that must be assembled into an integrated data base that can be accessed easily by the research teams for various analytical purposes. Therefore, one of the major tasks in Project A is to establish and maintain the longitudinal research data base (LRDB). This data base will link data on diverse measures gathered in the various tasks of Project A and, in addition, incorporate existing data that are routinely collected by the Army. Such a comprehensive LRDB will enable Project A to conduct a full analysis of how information gathered at each stage of the enlistee's progress through his/her Army career can add to the accuracy of predicting later performances.

The LRDB will not only facilitate efficient validation analyses that concern Project A, but it will also enable Project B to test and revise the prototype selection/allocation system. In addition, the usefulness of the LRDB extends beyond Projects A and B since it can also be used to support other research by the ARI staff (e.g., to address specific policy issues that might arise).

In planning for Project A, ARI anticipated the requirement for a longitudinal research data base to facilitate the storage and use of data throughout the life of the project. Recognizing the need for an early start, ARI staff began to collect, store, and edit a substantial amount of applicant, Military Applicant Profile (MAP), Enlisted Master File (EMF), training, and Skill Qualification Test (SQT) data for FY81, prior to Project A. These data served as the initial input to the LRDB.

Overview of LRDB Contents

In accordance with the Project A Research Plan, three major sets of data will be assembled within the LRDB. The first set consists of already existing data on FY81/82 accessions. These data include accession information (demographic/biographical data, test scores, and enlistment options), training success measures, measures of progress or attrition taken from the EMF, and specific information on SQT scores. This first set of data will be employed to validate the current version of the Armed Services Vocational Aptitude Battery (ASVAB) insofar as that can be done with available criteria. (This cohort was the first to receive forms 8/9/10 of the ASVAB.) It will also be used to investigate the major methodological and conceptual issues that must be resolved before the optimal estimates of the classification algorithm parameters can be made using the validation data from the 83/84 and 86/87 cohorts. (See Task 1 of the Project A Research Plan.)

The second and third major sets of data to be assembled into the LRDB will involve the new data collection efforts described in the research plan for the 83/84 and 86/87 cohorts.

General Objectives

The primary role of the LRDB is to support efficient data analyses as required by the research teams of both Projects A and B. The data collected throughout the research process of Project A and the data to be acquired from existing Army files must be organized and stored in such a way that they are simple and economical to access. Accordingly, the generation of the LRDB must meet the following objectives:

(1) It must develop systematic and efficient procedures for entering and editing the data.

(2) It must establish linkages of data from various sources and resolve all data inconsistencies.

(3) It must develop and maintain complete documentation of the data organization and contents.

(4) It must store both the data and the documentation cost-effectively and provide fast and easy access to both simultaneously.

(5) It must insure the security and integrity of the data.

Summary of First Year Activities

A significant portion of the first year's LRDB activities involved planning the data base contents and procedures for the duration of the project. The main result of this activity was the draft and final LRDB plan. Other planning accomplishments included the installation and testing of the RAPID data storage and retrieval system, the development of workfile generation and data set documentation programs, the identification and implementation of data file integrity and security procedures, and the evolution of data editing procedures.

Creation of the FY81/82 Cohort Data Base

Most of the substantive LRDB results during the first year were related to the creation of the FY81/82 cohort data base for use in the preliminary validation of the current ASVAB and the evaluation of new aptitude area composites. The FY81/82 cohort consists of 885,238 different individuals who applied for regular Army enlistment one or more times during FY81/82. The cohort includes a total of 268,297 regular Army accessions for whom subsequent progress and performance data have been assembled. Table 2 summarizes the types of data records that were assembled for the cohort and the number of applicants for which each type of record was found. A brief description of the different record types follows, along with a summary of the steps taken to enhance the accuracy and usability of the data received from each source.

Applicant/Accession Files. ARI receives a monthly data file from MEPCOM which contains information on all DoD applicants and accessions. The

Table 2

Number of Cases in the FY81/82 Cohort Data Base
by Data Source

DATA SOURCE	FY81	FY82	TOTAL
Applicant/Accession Cases			
Applicants not enlisted	273,175	264,839	538,103
Enlisted but not yet shipped	2,796	42,970	45,766
Discharged without being shipped	8,708	9,610	18,318
Accessions	136,928	128,794	265,722
Total Applicant Cases	421,607	446,212	867,819
MAP Cases			
Cases with MAP and accession data	4,618	3,794	8,412
Cases without applicant data			1,914
Total MAP Cases			10,326
Training Cases			
Cases with applicant data	49,728	6,077	55,805
Cases without applicant data			13,134
Total Training Records			68,939
EMF Cases			
Cases with applicant data	105,519	88,193	193,712
Cases without applicant data			2,575
Total EMF Cases			196,287
SQT Cases			
Cases with applicant data	47,746	12,167	59,913
Cases without applicant data			3,793
Total SQT Cases			63,706

regular Army applicants were selected out and created a merged file for each fiscal year. In the course of editing these data we

(1) "found" over 20,000 applicant records that had been inadvertently lost during the merging,

(2) identified over 5,000 cases with erroneous SSN codes (cases with identical names and birth dates and only one to three SSN digits mispunched or transposed),

(3) corrected over 2,000 cases where the ASVAB raw subtest scores had been misentered,

(4) unraveled the various uses of the different date fields, corrected errors, and created a set of date variables that are used consistently for all types of cases. (The date of the enlistment contract had been initially stored in the entry date field and then moved into the delayed entry program (DEP) date field at time of actual accession; we created a variable which was always the enlistment contract date and reserved the entry date field for only the actual accession date, for example, and made similar changes in instances where prior service cases had been treated differently from nonprior service cases),

(5) corrected inconsistencies in the recording of entry status,

(6) corrected, wherever possible, ubiquitous errors in the entry of the MOS fields, and in other key fields such as the ASVAB form code,

(7) resolved inconsistencies in sex, race, and enlistment program information between values on the accesion files and values on the Enlisted Masterfile, and

(8) began work on the documentation of these data that includes both codebooks giving the frequency of each value of each variable and a more detailed explanation of the meaning and use of each variable.

Some work remains to be done on the applicant/accession data including the editing of some variables not critical to the initial validation analyses (e.g., medical block data) and the resolution of additional interfile inconsistencies.

Military Applicant Profile (MAP) Data. Male applicants who are not high school graduates are required to complete a special biographical questionnaire from which an overall "suitability" score is derived. The questionnaire item responses are coded on scan sheets, but have been scored by hand. ARI has accumulated the scan sheets for most or all of the FY81/82 applicants who completed the MAP. During the first year, these sheets were scanned and the data loaded at NIH. Data on 10,326 different applicants

resulted from this activity. These data are now being checked, new scores are being generated, and attempts being made to resolve SSN errors that result in mismatches to the overall applicant files. When this editing is completed, the datafile documentation will be prepared and the data will be merged with the main applicant files.

Training Records Data. Training records were processed for 68,939 recruits who went through training during calendar year 1981. (Actually many more records were processed as some recruits went through more than one course.) As a first step, we reprocessed about 20,000 records for which significant information had not been entered initially. As with the applicant/accession data, the editing consisted of resolving erroneous or inconsistent values in every data field. The editing also included significant efforts to differentiate duplicate training records from cases where the same individual actually participated in more than one course. We also identified and corrected over 1,100 SSN errors that had led to a failure to match the training record to a corresponding applicant/accession record.

Enlisted Masterfile Data. Information from the FY82 year-end EMF was captured and entered into the LRDB. This information has been used to check key fields (e.g., race, sex, and SSN) in the other datafiles and to check against hardcopy records in the Task 4 effort to identify sources of information on general Army performance. These data will be used to assess soldier progress in the Army. Since these data are not involved in the initial validation analyses, the cleaning and documentation of this file has been given a lower priority and is only just now getting underway. Beginning with the last quarter in FY83, quarterly progress information will be extracted from the EMF for both the FY81/82 cohort and the FY83 cohort.

Skills Qualification Test Data. Data were received on SQT testing from FY79 through about January 1983. These files were found to contain SQT records for 59,913 of the individuals in the applicant file (some of whom were prior service cases and had taken an SQT prior to reentry into the Army in FY81/82). In general, only minor editing was required on the test date, MOS, and skill level fields.

Next Steps

Creation of a data base of this magnitude is a massive undertaking. Much progress has been made but it is not yet complete. During the next year the accumulation, editing, and documentation of the FY81/82 data files will be completed. The result will be the largest single data base ever created for purposes of personnel selection and classification research.

Also during this period: (a) additional special files will be created for specific research purposes, (b) data from the administration of the preliminary battery to the 83/84 longitudinal sample will be entered, (c) data from the pilot testing of training school measures and MOS specific measures will be entered in the data base, and (d) assistance will be provided to each of the other tasks and to ARI staff as they begin their initial analyses of Project A data.

Abstract

As noted in the Introduction, an abstract of a relevant research report follows.

LONGITUDINAL RESEARCH DATA BASE PLAN*

L. L. Wise and M. Wang
(AIR)
and
P. Rossmeissl
(ARI)

The research process of Project A will generate data to assemble into an integrated data base that can be accessed. Therefore, one of the major tasks in Project A is to establish and maintain the longitudinal research data base (LRDB). This data base will link together data on diverse measures gathered in the various tasks of Project A and, in addition, incorporate existing data that are routinely collected by the Army. Such a comprehensive LRDB will enable Project A to analyze fully how information gathered at each stage of the enlistee's progress through his/her Army career can add to the accuracy of predicting later performance.

*To be published as ARI Research Report 1356.

IV. VALIDITY ANALYSIS: PRELIMINARY ANALYSIS OF THE FY81/82 DATA FILE

The analyses of the FY 81/82 cohort data file will serve several purposes.

(1) The validity and differential validity of the existing predictors (ASVAB 8/9/10) against existing criteria (training grades, SQT, and administrative outcomes) will be determined on all MOS for which there are sufficient data. These results will serve as a benchmark against which the subsequent validations using new and/or improved predictors and criterion measures can be compared.

(2) The validity of alternative composites of ASVAB subtests can be compared with the validity of the existing composites.

(3) The validity generalization pertaining to both existing and alternative ASVAB composites can be modeled using the MOS clusters developed in Task 5.

(4) Alternative analytic methods for estimating the prediction parameters required by the classification system can be developed and evaluated using this data base (e.g., what method should be used to combine predictor information into a predicted score so as to maximize classification validity, minimize "shrinkage," and maximize robustness against cohort changes).

(5) The psychometric and distributional properties of the existing criterion measures can be determined so as to better describe their strengths and weaknesses.

The degree to which these objectives can be accomplished is a function of the size and completeness of the 81/82 cohort data file.

A summary of the initial steps to analyze the FY81/82 cohort data are described below.

Data Usability

The major purpose of the data usability analyses was to determine the degree to which the statistical assumptions underlying the validation analyses might be violated. The univariate distributions of the training and SQT criterion variables were obtained as well as the bivariate distributions of each criterion with each ASVAB subtest. Also, possible transformations of the criterion variables were evaluated using the same methods.

The findings for the univariate analyses are similar across MOS. In general, both criterion variables are negatively skewed (i.e., there are ceiling effects) with the SQT score typically less skewed than the corresponding training score.

Figure 4 illustrates a typical normal probability plot as obtained for MOS 94B with SQT score as criterion. (Note that if a distribution is, in fact, normal, the points should lie along a straight line running from the bottom left to the top right of the grid.)

Sample Sizes

A major concern for the 81/82 file data is that the proportion of the MOS in the enlisted occupational structure should contain sufficient cases to permit appropriate analyses. Therefore, we have attempted to include both training and SQT criterion measures so that separate validation analyses using each criteria can be compared. The training scores were collected as a part of ARI's preparation for this project, and we have been in the process of acquiring SQT data from Army computer files.

The current state of our data base for purposes of ASVAB validation, modeling classification and validity generalization, and determining criterion interrelationships is summarized in Table 3. The size of the sample for a given MOS was governed by the number of nonprior service enlistees who took one of the ASVAB 8/9/10 forms.

In preparing this table, we included MOS for which we have adequate data on at least 100 enlisted personnel, and we separated the data for different schools within the same MOS. Although the threshold of 100 is small for multivariate analyses, there is a trade off between sampling error and being able to include a sufficient number of MOS to model validity generalization. At N = 100, the standard error of the correlation coefficient is approximately .10.

Table 3 also indicates the MOS for which we have adequate data to perform subgroup validations. We tentatively consider availability of criterion scores for 50 or more enlistees within group as sufficient to support subgroup analysis. The lower threshold (50 instead of 100 previously used to determine data sufficiency for validation with the entire MOS) was adopted because we plan to employ simultaneous estimation technique to conduct the subgroup analysis. This approach uses both the within and between subgroup information to estimate subgroup parameters and thus tends to provide more stable estimates. Therefore, smaller sample sizes for each subgroup may be tolerated. However, we plan to emphasize only those subgroup analyses that are based on at least 100 cases. Of the 67 MOS with an adequate data base for the training criterion, 32 have sufficient numbers of both Blacks and Whites, 16 have sufficient numbers of both men and women, and one (76C) has sufficient numbers of all four combinations to support separate validation analyses. Of the 33 MOS with an adequate data base for the SQT criterion, 20 have sufficient numbers of both Blacks and Whites, 8 have sufficient numbers of both men and women, and two (05C and 94B) have sufficient numbers of all four combinations to support separate validation analyses.

It seems clear that we will not have sufficient data to examine the interactions between race/ethnicity and sex effects as regards the validity of all ASVAB subtests or composites. Also, due to sample size limitations, we cannot separately perform validation analyses for racial groups other than Blacks and Whites.

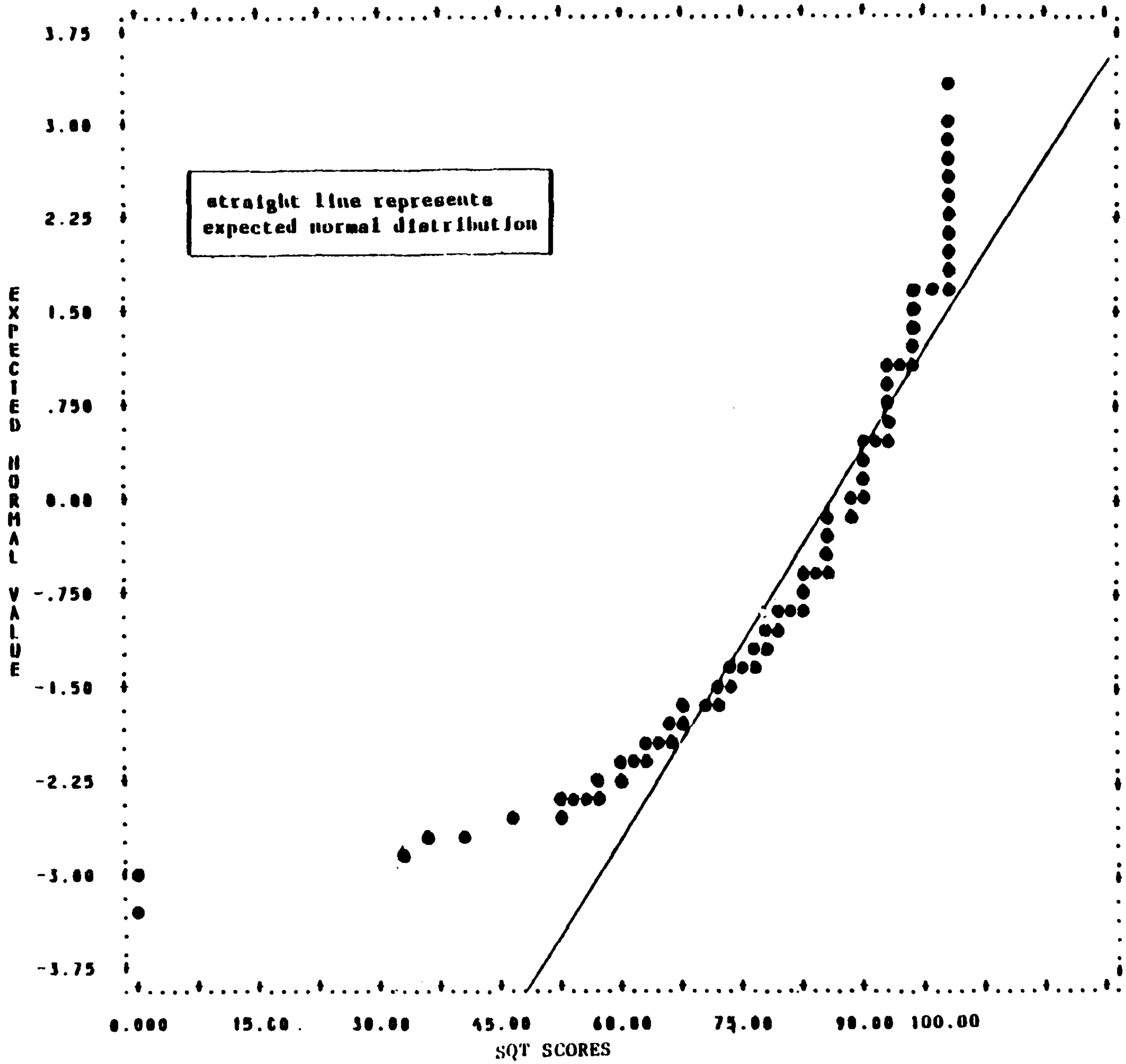


Figure 4. Normal Probability Plot of SQT Scores for FY81 Accessions in MOS 94B.

Table 3

Feasibility of Subgroup Validation and Total Sample Size for FY81 Accessions, by Prior Service and Criterion Availability

MOS	TCRS	AA	QS	GRPT	TOTTR	GRPS	TOTSQ
(Nonprior-Service, Both Training and SQT Criterion Available)							
05B	2A	SC	90	RACE AND SEX	517	RACE AND SEX	457
05C	2D	SC	95	RACE AND SEX	605	RACE BY SEX	889
11B	IN	CO	85	RACE ONLY	976	RACE ONLY	3430
11C	IN	CO	85	RACE ONLY	557	RACE ONLY	817
11H	IN	CO	85		428	RACE ONLY	572
12B	AB	CO	85		131	RACE ONLY	1015
12F	AF	CO	85	RACE ONLY	198		140
13B	3B	FA	85	RACE ONLY	639	RACE ONLY	1830
13E	3E	ST	95	RACE ONLY	449		280
13F	3F	FA	100	RACE ONLY	659	RACE ONLY	465
19D	9D	CO	85		186	RACE ONLY	527
19E	9E	CO	85		166	RACE ONLY	926
31M	4D	EL	95	RACE AND SEX	586	RACE AND SEX	653
31V	1V	EL	95	RACE ONLY	457	RACE ONLY	303
36C	AA	EL	90	RACE ONLY	179	RACE ONLY	195
36K	AC	EL	90	RACE ONLY	659	RACE ONLY	726
55B	5B	GM	85	RACE AND SEX	191	RACE ONLY	221
57H	G1	GM	85	RACE ONLY	169	RACE ONLY	123
62B	CB	MM	85		221		159
63B	3B	MM	85	RACE AND SEX	899		117
64C	EC	OF	85		199	RACE AND SEX	1403*
64C	4C	OF	85	RACE ONLY	398	RACE AND SEX	1403*
67N	65	MM	100		150		181
73C	5R	CL	95	RACE AND SEX	194	RACE AND SEX	242
75B	5E	CL	95	RACE ONLY	483	RACE ONLY	398
75D	5D	CL	95	RACE ONLY	228	RACE AND SEX	338
75E	5E	CL	95	RACE ONLY	268	RACE ONLY	165
76W	DB	CL	90		132		142*
76W	PW	CL	90	SEX ONLY	204		142*
82C	2C	ST	95		369		188
94B	KA	OF	85	RACE AND SEX	621	RACE BY SEX	1170*
94B	4B	OF	85	RACE AND SEX	625	RACE BY SEX	1170*
95B	SB	ST	100	SEX ONLY	716	RACE AND SEX	1370
(Nonprior-Service, Only SQT Criterion Available)							
12C	AC	MM	85		66		166
52D		GM	95		0		121
63B	DB	MM	85		3		117
75C	5D	CL	95		85		108

Table 3 (Cont'd)

Feasibility of Subgroup Validation and Total Sample Size for FY81 Accessions, by Prior Service and Criterion Availability

MOS	TCRS	AA	QS	GRPT	TOTTR	GRPS	TOTSQ
(Nonprior-Service, Only Training Criterion Available)							
110	[illegible]			RACE ONLY	326		0
15D	5D	OF	95		281		95
15E	5E	OF	95		264		74
16B	3A	OF	85		111		3
16E	EB	OF	95		117		0
16R	RA	OF	85	RACE ONLY	296		0
16S	SA	OF	85	RACE ONLY	506		1
17C	7C	SC	95	RACE AND SEX	189		71
17K	GA	EL	85		128		89
19F	9F	CO	85		123		6
27E	7E	EL	95		127		96
31N	4C	EL	95	RACE ONLY	159		84
32D	80	EL	95		116		1
45K	K9	GM	95		121		3
51K	3K	GM	85	RACE ONLY	155		0
54C	SS	GM	95		124		72
61C	HL	MM	100		123		0
63D	SA	MM	100		268		3
63G	M7	MM	100		100		0
63H	HL	MM	85		217		22
63N	TS	MM	95		274		10
63T	FI	MM	100		472		0
63W	WL	MM	85	RACE ONLY	276		4
63Y	TV	MM	100		130		8
67U	PL	MM	100		175		80
67V	18	MM	100		153		78
67Y	SL	MM	100		122		47
68J	W6	GM	95		102		34
68M	W8	GM	90		103		31
71N	L1	CL	95		117		68
76C	EC	CL	95	RACE BY SEX	1137		10
76P	SF	CL	90	RACE AND SEX	557		9
76V	EV	CL	90	RACE AND SEX	362		0
76Y	EY	CL	95	RACE AND SEX	377		10*
76Y	5G	CL	95	RACE AND SEX	297		10*
76Y	6Y	CL	95	RACE AND SEX	461		10*
91B	01	ST	95	RACE AND SEX	724		0
91C	02	ST	95	RACE AND SEX	220		0
91E	05	ST	95	SEX ONLY	154		5
92B	2S	ST	95	SEX ONLY	121		44

* Number of FY81 accessions having SQT scores for the MOS disregarding availability of training criterion for the individual; thus same for different courses of an MOS.

Note. AA = Current aptitude area composite;
QS = Current qualifying score;
TCRS = Training course;
GRPT = Possible subgroup validation with training criterion;
GRPS = Possible subgroup validation with SQT criterion;
TOTTR = Number of useable training records;
TOTSQ = Number of SQT scores available.

ASVAB Population Intercorrelations

Although criterion data are available only for recruits who actually enlist and are assigned to MOS, the selection and classification decisions must be applied to the entire population of applicants. To develop procedures when criterion data are missing for a large, nonrandom segment of the population, it is necessary to adjust for selectivity bias. A key component of this adjustment is the population covariance/correlation matrix of ASVAB subtests.

The report on the 1980 youth profile (18- to 23-year-olds) (Bock & Mislevy, 1981) provides one estimate for the potential applicant population. However, for purposes of this project, the ASVAB covariance/correlation matrix for FY81 nonprior service applicants has been estimated from a large sample (17,500) of the applicants taken from the total population of approximately 500,000 FY81 applicants. Table 4 represents means and standard deviations of ASVAB subtest scores for this sample. As shown, the applicants to the Army, on the average, score a half standard deviation below the norm (score of 50, see the unweighted means). The unweighted variance-covariance estimates are provided in Table 5a.

To compare the correlation matrix obtained for the FY81 Army applicants with that for the youth population, we weighted the sample to match the deciles of the AFQT. The weighted estimates are given in Table 5b, while the corresponding estimates from the 1980 youth population are shown in Table 5c. The weighted estimates are quite similar to the estimates from the youth population. The only difference between the two estimates that is larger than .10 is the correlation between PC (paragraph comprehension) and AS (auto/shop information), where the youth population estimate is .42 and the weighted estimate is .57.

The unweighted estimates of correlations are consistently lower than the corresponding weighted values. The average difference, however, is small (.04). We plan to use the unweighted estimates for the FY81 nonprior service applicant population to adjust for selectivity, because the present validation is specifically aimed to facilitate the selection and classification of applicants to the Army.

Pass/Fail Rates

The major criteria for the present validation are training success scores (e.g., end-of-course grade) and SQT score. In addition to the training course grade, the data also include indicators of end-of-course disposition: whether the soldier graduated and if not, the cause for nongraduation.

While the pass rates were generally quite high, there were significant differences in graduation rates among MOS/courses. For example, the graduation rates for Infantry (11B), Combat Engineer (12B), Cannon Crewman (13B), Motor Transport Operator (64C), and Utility Helicopter Repairer (67N) ranged from 91 percent to 100 percent; while for Short Range Gunnery Crewman (16R), Technical Engineering Supervisor (51K), Watercraft Engineer (61C), Aircraft Repairer (68J), and Personnel Actions Specialist (75E), the graduation rate ranged from 66 percent to 77 percent. Figure 5 highlights the differential graduation rates among the courses for these MOS.

Table 4

Summary Statistics for ASVAB Forms 8/9/10, Based on a 5% Sample of FY81 Nonprior-Service Applicants (Standard Scores are Used, N = 17,521)*

	ASVAB SUBTEST										
STAT	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI	VE
	UNWEIGHTED ESTIMATES										
MEAN	44.1	46.2	44.1	45.4	47.1	48.8	44.8	46.5	44.7	45.0	44.2
STD	10.1	9.1	10.4	10.3	10.5	9.4	10.0	8.5	9.1	9.3	10.3
	WEIGHTED ESTIMATES										
MEAN	48.9	51.3	49.3	50.2	50.9	51.8	48.2	50.9	48.4	48.6	49.5
STD	10.6	10.1	10.5	10.1	10.2	9.6	10.1	10.0	9.6	9.6	10.4

* Each subtest has a mean of 50 and standard deviation of 10 for the norming population.

Table 5a

Unweighted Estimates of Covariances and Correlations Among the Subtests of ASVAB Forms 8/9/10--Standard Scores, Based on a 5% Sample of FY81 Nonprior-Service Applicants*

(Above Diagonal = CORR, Below Diagonal = COV, Diagonal = VAR)

SUBTEST	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI	VE
GS	102.84	.68	.81	.71	.44	.37	.65	.61	.68	.70	.82
AR	62.88	83.57	.68	.65	.56	.45	.56	.75	.65	.61	.71
WK	85.25	65.08	108.19	.78	.50	.45	.60	.61	.[illegible]3	.68	.98
PC	74.34	61.64	84.31	106.83	.52	.47	.55	.58	.59	.61	.89
NO	46.26	53.63	53.97	56.23	109.37	.65	.30	.53	.37	.36	.53
CS	35.12	39.10	43.78	45.67	64.42	89.12	.25	.44	.33	.30	.47
AS	65.93	51.16	62.93	56.76	31.77	23.54	100.16	.44	.71	.74	.62
MK	52.67	58.31	53.88	51.04	47.18	35.41	37.62	72.00	.58	.53	.63
MC	62.27	54.28	59.52	55.35	35.42	28.08	64.83	44.51	82.17	.70	.65
EI	65.60	51.37	65.15	58.22	34.93	26.56	68.43	41.82	58.92	85.90	.69
VE	85.52	66.88	105.25	95.28	57.06	46.31	63.72	55.33	60.80	65.78	106.32

* The 5% sample includes 17,521 nonprior-service applicants who took ASVAB form 8/9/10. Because standard scores are used, each subtest has a variance of 100 for the norming population.

Table 5b

Weighted Estimates of Covariances and Correlations Among the Subtests of ASVAB Forms 8/9/10--Standard Scores, Based on a 5% Sample of FY81 Nonprior-Service Applicants*

(Above Diagonal = CORR, Below Diagonal = COV, Diagonal = VAR)

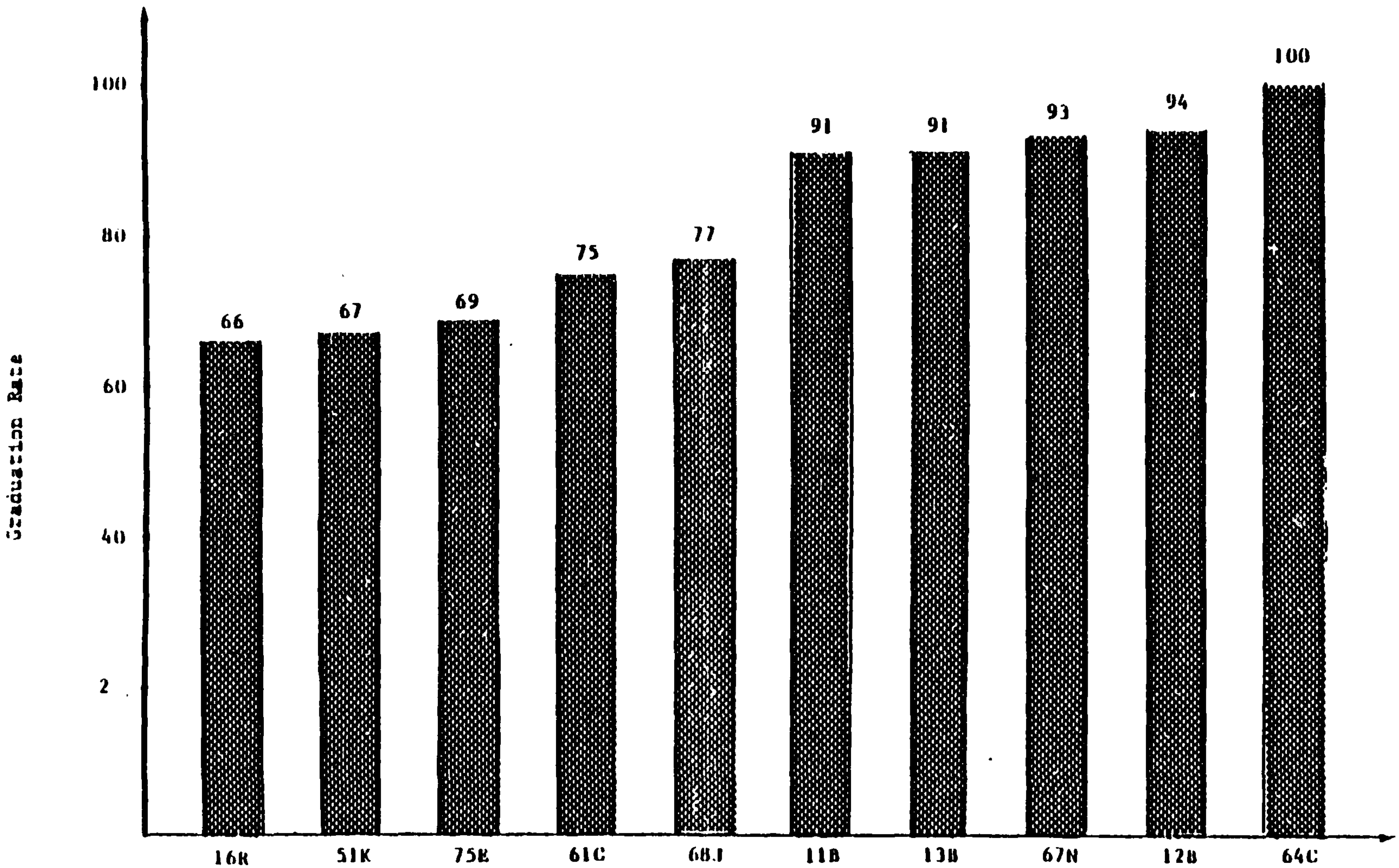
SUBTEST	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI	VE
GS	111.51	.74	.83	.75	.51	.42	.67	.68	.72	.74	.84
AR	78.39	102.00	.74	.71	.62	.52	.59	.81	.69	.65	.76
WK	92.46	78.33	110.23	.82	.57	.50	.61	.67	.65	.70	.98
PC	80.28	73.03	87.52	102.33	.58	.51	.57	.64	.62	.64	.91
NO	54.33	64.10	60.72	60.07	103.36	.68	.34	.58	.42	.41	.60
CS	42.96	50.23	50.07	49.40	65.85	91.75	.27	.50	.36	.34	.52
AS	71.43	60.04	64.83	58.21	35.28	26.20	102.81	.48	.73	.76	.62
MK	72.20	81.53	69.86	64.72	59.20	47.98	48.38	99.83	.63	.59	.68
MC	72.43	66.83	65.79	60.49	40.94	33.33	71.26	60.51	91.58	.74	.67
EI	75.24	63.30	70.37	62.27	40.13	31.01	74.11	56.31	67.89	92.60	.71
VE	92.63	80.09	107.75	96.26	63.16	52.07	65.59	71.23	67.00	70.86	108.95

* The 5% sample includes 17,521 nonprior-service applicants who took ASVAB form 8/9/10. Because standard scores are used, each subtest has a variance of 100 for the norming population.

Table 5c

Estimated Correlation Matrix of ASVAB Tests (Form 8/9/10), Based on 1980 Youth Population, 18-to-23 Years Old

TEST	GS	AR	WK	PC	NO	CS	AS	MK	MC	EI	VE
GS											
AR	72										
WK	80	71									
PC	69	67	80								
NO	52	63	60	60							
CS	45	51	55	56	70						
AS	64	53	53	42	30	22					
MK	69	83	67	64	62	52	41				
MC	70	69	60	52	40	34	74	60			
EI	76	66	68	57	41	34	75	59	74		
VE	80	73	98	90	62	57	52	70	60	67	



16R Short Range Gunner Crewman
51K Technical Engineer Supervisor
75E Personnel Actions Specialist
61C Watercraft Engineer
68J Aircraft Repairer

11B Infantry
13B Cannon Gunman
67N Utility Helicopter Repairer
12B Combat
64C Motor Transport Operator

Figure 5. Differential graduation rates among training courses for selected MOS (FY81 Accessions).

Some Preliminary Validation Results

To serve as examples of validation analyses, we have selected nine MOS representing each of the nine operational aptitude composites for preliminary analysis. These examples are based on samples of FY81 accessions who have both training and SQT scores available in the current data base.

Because the current data base does not include the scores for a large number of soldiers who have taken an SQT this year, the sample sizes are smaller than will be used in actual validation. (The SQT data for FY81 accessions will be increased substantially when the update tapes that ARI has requested arrive.)

Table 6 presents the summary statistics for ASVAB scores by MOS/course. The sample correlations of the subtest scores and the uncorrected simple validities (with training and SQT score as separate criterion) are shown in Table 7a. Tables 7b and 7c give the corrected subtest validities with training and SQT score as the criterion, respectively.

Next Steps

During the next contract period a number of important analyses will be carried out using the 81/82 data file. The most important of these follow:

(1) Validation of ASVAB subtests and current composites.

(2) Evaluation of alternative methods for adjusting selectivity bias (e.g., the design of paradigms that will allow the application of Heckman's procedure in the estimation of regressions from biased samples).

(3) Determination of appropriate number of composites and forming homogeneous MOS groups for the development of new composites.

(4) Evaluation of the discriminant validity of the new composites employing optimal assignment algorithms.

(5) Investigation of moderator effects and differential validities among subgroups of enlisted personnel.

(6) Cross-validation of the newly developed composites.

(7) Modeling the generalization of validation results to the population of Army jobs.

Each of these activities will produce a technical report. A final report on the recommendation for new composites and qualification scores for ASVAB forms 11/12/13 will also be prepared.

Table 6

Summary Statistics of ASVAB and Criterion Scores for FY81 Accessions in Nine MOS (Sample includes only soldiers with both training and SQT scores)

MOS	TSCHL	TCRS	STAT	GS	AR	WK	PC	VE	NO	CS
05C	113	2D	MEAN	51.1	53.3	53.2	54.0	53.4	55.7	55.8
05C	113	2D	STD	7.9	8.1	6.1	5.8	6.1	6.1	6.8
11B	809	IN	MEAN	50.1	51.3	50.4	51.6	50.8	50.6	51.5
11B	809	IN	STD	8.3	7.6	7.6	7.4	7.2	7.6	6.4
13B	810	3B	MEAN	44.2	48.8	44.6	46.7	44.9	50.3	50.8
13B	810	3B	STD	9.4	7.3	8.8	8.5	8.6	6.8	5.9
31M	113	4D	MEAN	51.6	53.4	51.2	52.5	51.7	50.5	51.3
31M	113	4D	STD	6.2	6.4	6.8	6.4	6.4	8.4	7.6
55B	093	5B	MEAN	47.1	43.0	46.9	47.2	46.6	43.8	46.2
55B	093	5B	STD	6.2	6.9	6.4	7.2	6.3	7.0	7.2
62B	807	CB	MEAN	47.6	48.4	46.9	47.9	46.6	49.2	49.6
62B	807	CB	STD	8.1	7.9	8.0	8.2	8.1	7.6	7.5
75B	121	5E	MEAN	45.7	49.3	48.2	49.3	48.5	56.3	56.4
75B	121	5E	STD	8.7	8.5	8.2	8.3	8.0	5.1	6.0
94B	101	KA	MEAN	47.0	48.7	48.7	50.4	49.0	51.1	49.6
94B	101	KA	STD	7.4	7.3	6.6	6.4	6.2	7.3	8.3
95B	813	SB	MEAN	54.9	55.1	55.5	55.4	55.7	53.5	53.7
95B	813	SB	STD	5.7	7.3	4.9	5.1	4.4	7.4	8.3

MOS	TSCHL	TCRS	STAT	AS	MK	MC	EI	TSCR	SQT	N
05C	113	2D	MEAN	51.4	52.1	50.0	50.5	88.9	76.2	343
05C	113	2D	STD	8.3	8.3	8.1	7.5	6.0	9.8	
11B	809	IN	MEAN	51.9	49.5	50.9	50.5	94.8	87.0	575
11B	809	IN	STD	7.5	8.1	7.2	7.5	5.0	7.6	
13B	810	3B	MEAN	43.7	48.3	46.3	45.2	78.5	86.5	374
13B	810	3B	STD	10.1	6.7	7.4	8.6	18.5	9.1	
31M	113	4D	MEAN	48.2	51.8	48.8	51.3	93.0	86.6	272
31M	113	4D	STD	8.7	6.4	6.8	6.7	4.9	9.9	
55B	093	5B	MEAN	45.1	45.3	41.5	47.0	85.8	79.1	100
55B	093	5B	STD	6.0	5.6	6.1	5.0	4.6	10.0	
62B	807	CB	MEAN	53.8	47.1	51.4	50.9	91.5	79.8	121
62B	807	CB	STD	7.3	7.1	7.8	7.4	6.7	9.4	
75B	121	5E	MEAN	44.5	49.6	44.7	46.3	86.5	66.9	263
75B	121	5E	STD	8.7	7.6	8.0	8.0	11.6	17.8	
94B	101	KA	MEAN	48.1	47.2	48.4	46.9	86.6	87.5	320
94B	101	KA	STD	7.5	6.8	7.0	7.7	1.6	9.0	
95B	913	SB	MEAN	54.0	53.5	53.6	52.8	81.3	85.9	449
95B	813	SB	STD	7.4	7.5	6.4	6.9	6.6	9.2	

Note. TSCHL = Training school code; TCRS = Training course code; ASVAB subtest scores are standardized so that the mean for the norm population is 50 and standard deviation is 10.

Table 7a

Sample (Uncorrected) Validity Coefficients of ASVAB Tests by MOS and Course

GROUP	COURSE	N	CRIT	GS	AR	VE	NO	CS	AS	MK	MC	EI	TSCR	SQT
05C:	113: 2D	243	TSCR	.18	.26	.15	-.01	.02	.14	.22	.23	.20		.10
05C:	113: 2D		SQT	.26	.29	.21	-.06	-.05	.25	.26	.26	.28	.10	
05C: W	113: 2D	239	TSCR	.13	.23	.15	.03	.13	.10	.19	.19	.16		.13
05C: W	113: 2D		SQT	.17	.25	.22	-.01	.01	.19	.24	.23	.21	.13	
05C: B	113: 2D	86	TSCR	.13	.16	.09	-.00	-.11	.07	.18	.18	.09		-.07
05C: B	113: 2D		SQT	.13	.13	.06	.01	-.08	.06	.16	.05	.18	-.07	
05C: F	113: 2D	58	TSCR	.13	.25	.18	.18	-.03	.24	.15	.18	.30		.10
05C: F	113: 2D		SQT	.01	.24	.10	.05	.13	.03	.26	.02	.06	.10	
05C: M	113: 2D	285	TSCR	.19	.26	.14	-.04	.04	.13	.23	.24	.18		.10
05C: M	113: 2D		SQT	.29	.27	.23	-.04	-.06	.24	.24	.27	.30	.10	
11B:	809: IN	575	TSCR	.22	.24	.19	.12	.10	.18	.27	.27	.15		.16
11B:	809: IN		SQT	.25	.27	.27	.11	.07	.23	.30	.25	.26	.16	
13B:	810: 3B	374	TSCR	.10	.12	.13	.07	.01	.15	.04	.15	.18		.11
13B:	810: 3B		SQT	.29	.25	.29	.12	.07	.32	.19	.29	.22	.11	
31M:	113: 4D	272	TSCR	.12	.26	.16	.15	.17	.01	.25	.12	.08		.23
31M:	113: 4D		SQT	.19	.14	.09	-.03	-.04	.07	.15	.11	.16	.23	
31M: W	113: 4D	171	TSCR	.22	.29	.18	.19	.17	-.04	.29	.11	.20		.27
31M: W	113: 4D		SQT	.22	.16	-.00	-.08	-.13	.01	.17	.08	.11	.27	
31M: B	113: 4D	87	TSCR	-.20	.13	-.00	.09	.13	-.19	.22	-.01	-.18		.20
31M: B	113: 4D		SQT	.14	.13	.27	.09	.15	.13	.17	.13	.20	.20	
31M: F	113: 4D	76	TSCR	.14	.35	.06	.13	.25	.03	.16	.05	.12		.14
31M: F	113: 4D		SQT	.14	.21	-.10	-.16	-.05	.18	.01	.10	.05	.14	
31M: M	113: 4D	196	TSCR	.14	.25	.16	.12	.09	.07	.26	.22	.14		.25
31M: M	113: 4D		SQT	.22	.12	.14	-.01	-.06	.07	.19	.13	.22	.25	
55B:	093: 5B	100	TSCR	.22	.28	.21	.04	.11	.20	.19	.07	.15		.29
55B:	093: 5B		SQT	.17	.19	.12	.17	.13	.06	.35	.07	-.00	.29	
62B:	807: CB	121	TSCR	.32	.29	.27	-.03	.27	.42	.29	.32	.21		.22
62B:	807: CB		SQT	.36	.30	.33	-.04	.13	.39	.31	.39	.19	.22	
75B:	121: 5E	263	TSCR	.17	.34	.17	.04	.01	.23	.27	.17	.15		.34
75B:	121: 5E		SQT	.35	.45	.37	.06	.05	.27	.38	.34	.26	.34	
94B:	101: KA	320	TSCR	.13	.18	.15	.02	.11	.19	.18	.07	.18		.11
94B:	101: KA		SQT	.18	.19	.17	-.16	.02	.18	.12	.19	.17	.11	
94B: W	101: KA	204	TSCR	.09	.13	.14	.13	.13	.11	.13	-.03	.12		.04
94B: W	101: KA		SQT	.17	.12	.18	-.20	-.02	.15	.11	.15	.14	.04	
94B: B	101: KA	109	TSCR	.05	.11	.05	-.16	.01	.19	.20	.13	.19		.26
94B: B	101: KA		SQT	.01	.22	.04	-.04	.02	.08	-.02	.13	.10	.26	
94B: F	101: KA	60	TSCR	.23	.21	.29	-.14	.19	.37	.30	.33	.19		.29
94B: F	101: KA		SQT	.18	.04	.20	-.12	-.06	.04	.06	.08	.07	.29	
94B: M	101: KA	260	TSCR	.13	.18	.08	.02	.02	.28	.17	.10	.27		.09
94B: M	101: KA		SQT	.18	.21	.17	-.17	.04	.22	.13	.22	.19	.09	
95B:	813: SB	449	TSCR	.30	.22	.30	.07	.10	.21	.16	.21	.32		.14
95B:	813: SB		SQT	.17	.24	.15	.15	.11	.22	.23	.24	.23	.14	

Note. GROUP = MOS name followed by subgroup identification (W for White, B for Black; and F for Female, M for Male).
TSCR = Training course grade. SQT = SQT percentage score.
The last two columns provide the intercorrelations between the training and SQT criterion scores.

Table 7h

Corrected Validity Coefficients of ASVAB Tests (Form 8/9/10) by MOS Groups, Criterion is Training Score (Corrections based on unweighted covariance matrix for FY81 nonprior-service applicants)

MOS_RS	N	GS	AR	VE	NO	CS	AS	MK	MC	EI
05C:	343	.26	.31	.26	.19	.20	.23	.27	.29	.26
05C: W	239	.29	.35	.33	.25	.31	.27	.30	.32	.30
05C: B	86	.15	.16	.10	.05	-.00	.12	.18	.19	.13
05C: F	58	.51	.55	.56	.64	.42	.59	.49	.53	.60
05C: M	285	.21	.25	.18	.10	.15	.17	.23	.25	.19
11B:	575	.32	.35	.29	.24	.23	.29	.34	.36	.25
13B:	374	.12	.14	.15	.14	.06	.19	.08	.19	.22
31M:	272	.46	.51	.48	.37	.35	.31	.48	.40	.41
31M: W	171	.55	.57	.54	.44	.39	.35	.53	.42	.54
31M: B	87	-.15	-.03	-.09	-.01	.06	-.22	.06	-.08	-.16
31M: F	76	.45	.57	.38	.38	.35	.30	.47	.31	.39
31M: M	196	.46	.51	.47	.32	.26	.36	.47	.46	.44
55B:	100	.61	.57	.57	.27	.24	.54	.54	.49	.60
62B:	121	.43	.40	.41	.27	.36	.51	.38	.43	.40
75B:	263	.23	.33	.22	.13	.13	.28	.28	.23	.24
94B:	320	.30	.34	.35	.24	.26	.34	.32	.27	.34
94B: W	204	.32	.35	.39	.34	.33	.33	.31	.24	.33
94B: B	109	.34	.38	.36	.17	.23	.41	.41	.39	.41
94B: F	60	.45	.50	.56	.21	.28	.60	.52	.59	.56
94B: M	260	.27	.30	.25	.18	.14	.38	.26	.27	.37
95B:	449	.59	.52	.61	.33	.32	.48	.45	.49	.57

Note. MOS_RS = MOS name followed by subgroup identification (W for White, B for Black; and F for female, M for male).

Table 7c

Corrected Validity Coefficients of ASVAB Tests (Form 8/9/10) by MOS Groups, Criterion is SQT Score (Corrections based on unweighted covariance matrix for FY81 nonprior-service applicants)

MOS_RS	N	GS	AR	VE	NO	CS	AS	MK	MC	EI
05C:	343	.36	.37	.36	.23	.19	.36	.34	.34	.38
05C: W	239	.34	.38	.40	.26	.22	.37	.35	.35	.37
05C: B	86	.20	.17	.15	.11	.05	.16	.21	.13	.26
05C: F	58	.06	.24	.16	.04	.20	.08	.27	.07	.12
05C: M	285	.42	.39	.42	.25	.19	.40	.35	.39	.43
11B:	575	.36	.38	.38	.26	.21	.34	.37	.36	.36
13B:	374	.34	.33	.35	.25	.19	.35	.29	.34	.27
31M:	272	.42	.39	.36	.19	.13	.30	.38	.32	.38
31M: W	171	.31	.27	.19	.05	-.01	.17	.26	.18	.23
31M: B	87	.58	.58	.62	.42	.37	.47	.55	.52	.58
31M: F	76	.12	.20	-.10	-.15	-.16	.23	.03	.12	.10
31M: M	196	.45	.41	.40	.18	.10	.35	.40	.39	.46
55B:	100	.50	.49	.47	.32	.28	.40	.59	.42	.44
62B:	121	.48	.44	.48	.29	.26	.49	.42	.51	.40
75B:	263	.54	.61	.60	.48	.45	.39	.55	.49	.45
94B:	320	.24	.25	.25	.01	.08	.26	.18	.25	.24
94B: W	204	.19	.17	.17	-.10	-.00	.18	.15	.17	.17
94B: B	109	.44	.52	.51	.37	.31	.48	.34	.51	.48
94B: F	60	.33	.21	.33	.09	.06	.21	.22	.25	.21
94B: M	260	.24	.26	.23	.00	.08	.28	.17	.27	.25
95B:	449	.40	.43	.41	.33	.27	.42	.41	.44	.43

Note. MOS_RS = MOS name followed by subgroup identification (W for White, B for Black; and F for female, M for male).

Abstracts

As noted in the Introduction, abstracts of relevant and related research reports follow.

VALIDITY OF ASVAB 8/9/10 FOR PREDICTING TRAINING SUCCESS*

P. G. Rossmeissl, C. J. Martin, H. Wing
(ARI)
and
M. Wang
(AIR)

Like the rest of the armed services, the Army makes use of aptitude area composites formed from the 10 ASVAB subtests to select and classify potential enlisted personnel. This paper reports upon the validity of these composites as predictors of Army training success. Eleven different military occupation specialities (MOS) were included in the research because previous work had collected useful training criterion data for each MOS. Three sets of composites, including AFQT, operational Army composite, and the newly proposed high school composites were validated against these criteria. The results showed that each type of composite was highly predictive of training success within the Army. A further detailed analysis of those MOS in the sample with large values of N showed that validity of the composites did not vary significantly as a function of race or gender.

FACTORIAL INVARIANCE OF THE ARMED SERVICES* VOCATIONAL APTITUDE BATTERY

L. M. Hanser and K. J. Mitchell
(ARI)

The purpose of this research was to examine the factoral invariance of the Armed Services Vocational Aptitude Battery (ASVAB) across sex and racial/ethnic population subgroups. Samples included 79,926 males, 18,728 females, 62,389 Whites, 29,546 Blacks, and 656 Hispanics. Three hypotheses were sequentially tested for each subgroup using the LISREL V program developed by Joreskog. The three hypotheses were: (1) subgroup covariance matrix were equal, (2) the numbers of factors were invariant across subgroups, and (3) factor structures were similar across subgroups.

Analyses were directed by identifying possible sources of differential predictive validity for sex and racial/ethnic subgroups, such as differences in measures constructs or sources of measurement error. Results speak to differential prediction of training/job performance for population subgroups and to the efficacy of employing a common composite system for the selection and classification of all applicant groups.

---

*Paper presented at the 25th Annual Conference of the Military Testing Association in Gulf Shores, Alabama, October 1983.

EVALUATION OF THE ASVAB 8/9/10 CLERICAL (CL) COMPOSITE* FOR PREDICTING TRAINING SCHOOL PERFORMANCE

M. Weltin and B. A. Popelka
(ARI)

The composite of Armed Services Vocational Aptitude Battery (ASVAB) subtests used to select applicants for entry-level training in Army clerical schools was evaluated by correlating composite scores with training performance scores. The clerical composite (CL) had high validity ($r=.68$) for this criterion, but an alternate composite of arithmetic reasoning, paragraph comprehension, and mathematics knowledge scores provided from multiple regression analyses had even higher validity ($r=.74$). Differential prediction for classification purposes is discussed.

VALIDITY OF THE MILITARY APPLICANT PROFILE FOR PREDICTING EARLY ATTRITION IN DIFFERENT EDUCATIONAL, AGE, AND RACIAL GROUPS**

N. K. Eaton, M. Weltin and H. Wing
(ARI)

The Military Applicant Profile (MAP) was developed to serve as an applicant screening instrument to reduce attrition. Since 1979, it has been used operationally to screen 17-year-old non-high school graduate males. The Army Research Institute (ARI) was asked to explore the extension of MAP to older (above 17) nongraduate males. Using 1976-77 data, this research evaluated the validity of MAP for education, race, and age subgroups. Results showed that MAP scores were significantly related to the 180-day stay-leave attrition criterion. Neither race (Black-White) nor age interacted with the MAP-attrition relationship; education level did. The function relating MAP scores to attrition for graduates was significantly below that for nongraduates. MAP would appear to have great utility in reducing the much higher attrition rate of nongraduates. Research to verify these findings is currently underway.

---

*To be published as ARI Technical Report 594.
**Published as ARI Technical Report 567.

READING ASSESSMENT IN THE ARMY*

R. L. Oxford-Carpenter and L. J. Schultz
(ARI)

This report describes research in the area of reading assessment in the U.S. Army. The purpose of the research is to discover the best ways for the Army to use reading assessment in order to increase productivity and effectiveness in training and on the job.

The research introduces a cognitive theory of reading assessment based largely on information processing models. Extensive discussion on the differences between aptitude and achievement tests leads to an explanation of types and purposes of reading tests. Examples include norm-referenced reading tests for ranking individuals, objective-referenced reading tests for determining how an individual performs relative to a set of objectives, and criterion-referenced reading tests for deciding whether an individual has met specified standards. Crucial characteristics of reading tests that must be considered in test selection are examined in the report: main test purposes, intended population, reliability, validity, appropriate norms, available scores, administrative ease, multiple forms availability, item quality, and relevance to Army needs. Test bias--sometimes known as differential validity--is a key topic.

---

*To be published as Selection and Classification Technical Area Working Paper 83-7.

V. CRITERION DEVELOPMENT

A large proportion of our efforts during the first two years of Project A is being devoted to the development of criterion measures for assessing training and job performance. Estimating the prediction parameters for a selection and classification system that must place so many people in such a variety of jobs demands the most complete and precise information that we can gather. Consequently, the validation of predictors must be based on reliable, meaningful, and comprehensive criteria. To the extent that the criteria on which our statistical estimates are based lack relevance and are unreliable or deficient, the effectiveness of the classification system will suffer.

Rather than simply pick whatever traditional criterion measures happen to be available we have elected to conceptualize the criterion problem in construct validity terms. The strategy is to begin with a conceptual model of the entire work performance environment, incorporating external and organizational influences on performance, the person and job components of behavior, as well as their interaction, organizational controls which impact on performance measurement, and performance outcomes. An abstract of a paper describing this model appears at the end of this section. This broad integrative model will provide a context in which a description of the criterion space can be developed that we believe will account for a large proportion of the major facets of soldier performance and effectiveness. We will proceed from this specification of criterion constructs to the development of an integrated set of criterion measures that reflect individual task proficiency, contributions to general organizational goals, minimization of human-resource-related costs, and the relative utility of performance across jobs.

Ideally, if our criterion development efforts were to proceed according to the latest technical thinking as regards conceptualizing a domain of latent variables or constructs, we would most likely adopt a structural model as the guiding heuristic (e.g., James, Muliak, & Brett, 1982). That is, the initial focus would be on trying to specify, however imperfectly, the latent variables or constructs that comprise the criterion space, as well as the nature and degree of their interrelationships. The next step would attempt to specify the manifest, or measureable, variables that represent each latent variable and to predict how the manifest variables are interrelated. The relevant issues then become:

(1) How good is our current theory and knowledge about each latent variable and about how they should interrelate (causally or otherwise)? Unfortunately, applied psychology in general knows a lot more about the latent structure of the predictor side than the criterion side.

(2) Are all the relevant latent variables measured by one or more manifest variables? Is there redundancy? Are some constructs unmeasured?

(3) How much do we know about the validity of the manifest variables as a measure of the latent constructs? How much more do we need to find out?

(4) Should two manifest variables be related to one another? If so, is it because they are measures of the two constructs that stand in a causal relation to one another?

(5) Can we specify other factors that will determine the relationship between two manifest variables besides their relationship to the latent structure (e.g., unreliability, common method variance, "halo," the implicit theories of performance held by raters, etc.)?

These are not easy questions to answer but they are relevant to building an understanding of criteria and their interrelationships. In the best of all possible worlds, the explication of such a structural model will be an iterative process over the course of such a large project. Although it may never get to a statistically "testable" form in the confirmatory analysis sense, it would be refined on the basis of each new increment of research data, and it would also guide data collection and analysis. Consequently, it is very much a bootstrapping process. Its aim is to maximize our understanding of the criterion space as well as to provide a basis for developing a composite criterion for validation purposes.

The project began with the basic premise that there are three major components to the total criterion (job effectiveness) space: (a) the individual's performance and effectiveness during training, (b) performance on the specific job tasks for which the individual is responsible, and (c) aspects of performance and effectiveness that are not MOS specific but that are a major part of the effectiveness of every enlisted person. Within each of these major domains the task of the project is to explicate the constructs that define it, develop operational measures of these constructs, and combine operational measures into criterion composites that are maximally useful for developing the selection and classification system.

Relative to the above framework, work on each of the three major criterion domains began in earnest during the past eight months. These efforts are described in the following chapters.

Abstract

As noted in the Introduction, an abstract of a relevant and related research report follows.

JOB PERFORMANCE AND ASSESSMENT:
A SYSTEMIC MODEL*

L. I. Wetrogan, D. Olson, and H. Sperling
(ARI)

The Army has recently instituted a multiyear, large-scale research effort to improve its soldier selection and classification system. A major objective of this effort is the development of performance criteria which can be used in evaluating the validity of current and future predictor measures. The successful achievement of this objective will depend upon the manner in which problems that have historically made performance measurement a particularly challenging undertaking are addressed. The present paper discusses a conceptual model which has been constructed to guide research on soldier performance.

The model conceptualizes work performance in the Army as a complex multidimensional process which is determined by a diverse group of individual, job, environmental, and organizational factors. Thus, a comprehensive and fluid system framework has been developed to define relevant performance variables and to examine methodologies that could be used to assess performance dimensions in relation to various designated criteria. It is expected that the model will assist in understanding which patterns of individual differences and contextual situational factors contribute to observed performance differences, that it will provide a basis for identifying important research areas, and that it will serve as a tool for explaining research findings.

In accordance with a systems approach to work performance, this model building endeavor has identified the following component subsystems: (1) organization, (2) person, (3) job, (4) environment, (5) work performance, and (6) performance evaluation system. The model describes these subsystems and their interdependent relationships.

*Paper presented at the 1983 Annual Convention of the American Psychological Association in Anaheim, California, August 1983.

VI. MOS TASK DESCRIPTIONS

Several initial phases of Project A criterion development activities are dependent on: (a) the generation of a set of task descriptions for each focal MOS and (b) the identification of job behaviors which are not specific to a particular MOS but which are critical for effective performance in general. That is, we need job descriptions that are comprehensive in terms of covering performance factors that are both common to all MOS and MOS specific.

Two different methods are being used to provide this pool of information. One makes principal use of the Army's existing sources of task descriptions for each MOS. The second uses the critical incident method to generate a large pool of critical performance behaviors for specific MOS and that cut across MOS. The two methods are being used for somewhat different purposes and each is discussed below.

MOS-Specific Task Descriptions

The task descriptor "item banks" are being used in the following ways:

(1) To construct knowledge tests for the assessment of training achievement, a sample of task descriptors was drawn for each MOS stratified by task duty area and frequency. These tasks were then judged in terms of their match with training objectives and current doctrine regarding job content. The task descriptions and their associated judgments will then serve as one basis for the generation of knowledge test items.

(2) The task descriptor item bank is also the principal starting point for the construction of hands-on performance measures to be described in a later section of this report. Several procedural steps are being used to select tasks from this pool so that the resulting sample of tasks is the most appropriate for standardization as performance test items.

(3) A computerized content analysis procedure has been applied to the task item banks as a check against whether important performance factors are being missed as the various steps in our criterion development efforts proceed.

(4) Later in the project, the task descriptions will be used to develop a common set of descriptors that can be used to describe all MOS and cluster them empirically into homogeneous job families.

For Project A purposes two sources of job analytic information were consolidated for each selected MOS at Skill Level 1. The first source was the Soldier's Manual (SM); the second was Comprehensive Occupational Data

Analysis Program (CODAP) frequency data from the Army Occupational Survey Program (AOSP).

The SM specifies the tasks that by doctrine are critical to the soldier's job performance at a given skill level (SL). These critical tasks represent a subset of the tasks a soldier could perform. The immediate purpose of the SM is to guide training on the critical subset. Although the procedure to identify the critical subset for a particular MOS varies by proponent (i.e., the unit responsible for training in a particular MOS), task selection for the SM is typically a high visibility activity that involves the highest levels of a proponent's command. In addition to the tasks in the MOS SM, 71 common tasks have been designated as critical to the job performance of all SL1 soldiers. These tasks are listed in the Soldier's Manual of Common Tasks (FM21-1).

As just noted, CODAP is a description of job activities based on a checklist survey of job incumbents. The checklist contains items describing a variety of duties and tasks related to the MOS. The items have been drawn from job analysis materials and subject matter experts. Although the items are intended to reflect job content, by virtue of the way they are generated the items reflect the intended content of the MOS as well as the actual content.

The consolidation of SM and CODAP served three purposes for selecting task descriptions to serve as a basis for criterion development:

(1) Describes domain of the soldier's job

(2) Determines frequency for critical tasks

(3) Confirms completeness of SM.

Describes Domain of the Soldier's Job. A soldier's job consists of tasks and activities he or she is trained to perform (doctrine) at a particular skill level, and tasks and activities that he or she actually performs on the job. Although there really is a major overlap, differences exist between the two sets. For example, in field environments there is seldom the sharp distinction between different skill levels that exists doctrinally. Job doctrine is best reflected in SM while CODAP generally provides a fuller picture of field requirements. Integrating the two documents gives a more complete view of the domain that will be experienced by the SL1 soldier.

Determines Frequency for Critical Tasks. The Soldier's Manual does not provide data on which tasks are most widely performed within a skill level; however, that information is available in CODAP. Having that information protects against randomly selecting, for example, a 13B SM task like "operate intercommunications systems," which is apparently performed by only three percent of cannoneers.

Confirm Completeness of SM. Checking the job descriptions from CODAP against job descriptions from the SM also insured that potentially critical

tasks were not lost. For example, during the transition to centralized common task management the task "engage targets with an M16" was not included in any of the MOS SM or in the Soldier's Manual of Common Tasks. There also may have been shifts in analysts' assumptions about the scope of tasks which, in conjunction with changes in the MOS, have caused potential critical tasks to be overlooked. For example, the 13B activity "clean cannon tube and chamber" may have at one time been assumed to be part of preventive maintenance checks and services (PMCS). It does not now appear to be part of PMCS or any other SM task even though 74 percent of cannoneers report doing it.

Critical Incident Descriptions

The job behavior descriptions being generated by the critical incident method serve the following purposes:

(1) They are a major source of information for the identification and explication of the factors that define job-performance and effectiveness, both general and MOS specific.

(2) They are the primary means by which rating scale measures of general and MOS-specific performance factors will be constructed.

The critical incident procedure involves the following general steps:

(1) Workshops comprised of 10-20 supervisors (NCO and/or officers) are asked to generate specific examples of job behaviors for enlisted personnel that are "critical" in terms of reflecting positive or negative aspects of performance.

(2) The specific incident descriptions are then categorized by a panel of judges into categories that seem to reflect the major underlying performance factors.

(3) Another group of judges then "retranslates" the specific incidents by assigning them to the performance category in which they best fit. To the extent that this retranslation can be done reliably the category system is a meaningful one.

(4) At the same time the incidents are retranslated they are also judged, or scaled, in terms of the level of effective or ineffective performance they represent.

VII. DEVELOPMENT OF TRAINING MEASURES

General Purpose

The general purpose of the research on training criteria is to generate information about training performance which can be used in the validation of initial predictors and in the prediction of first-tour and second-tour performance in the Army. To accomplish this purpose, existing measures of training performance are being analyzed and evaluated, new measures are being developed where needed, and composite sets of predictor and criterion measures will be assembled. As job performance surrogates, training measures can serve to reduce the time required for predictor validations from years to months. When used to predict subsequent performance, training measures have the potential to increase the accuracy of classification into MOS over that obtained by the use of preinduction predictors alone. Both the extent to which training measures can be used as surrogates for ultimate job performance criteria and the degree of incremental validity obtained by including training success as a predictor itself will be assessed during the course of the project.

First Year Activities

The project activities during the first year of training criterion development have concentrated on (a) a review of the literature, (b) the analysis of current school measures, (c) the documentation and analysis of training objectives and training content, and (d) the first steps in the development of comprehensive job knowledge tests. The development of training measures is being accomplished for 19 selected MOS.

Analysis of Current School Measures

Preliminary identification of current school measures as criteria of training performance and as predictors of subsequent job performance is being accomplished through analyses of test score distributions, a review of the test construction and test scoring process, and a comparison of the measures' coverage of training and job content.

To date, instructors and supervisors in the following 11 courses have been interviewed:

05C - Radio Teletype Operator	Ft. Gordon, GA
16S - Manpads Crewman	Ft. Bliss, TX
19E - Tank Crewman	Ft. Knox, KY
19K - M-1 Crewman	Ft. Knox, KY
63B - Light-Wheeled Vehicle Mechanic	Ft. Dix, NJ
63B - Light-Wheeled Vehicle Mechanic	Ft. Jackson, SC
64C - Motor Transport Operator	Ft. Dix, NJ
71L - Administrative Specialist	Ft. Jackson, SC

76Y - Unit Supply Specialist | Ft. Jackson, SC
94B - Food Service Specialist | Ft. Dix, NJ
94B - Food Service Specialist | Ft. Jackson, SC

The interview was concerned primarily with trainee progress and achievement measures in each course. There was surprising unanimity among the courses in these matters. All of the courses are group-paced (GP), except for the self-paced (SP) 05C course, the mostly self-paced 16S course, and the lock-step (LS) 19E and 19K courses. Since group-paced and lock-step are virtually indistinguishable modes of procedure, the 05C and 16S courses are the only real exceptions and both are scheduled to become group-paced in the near future.

Training performance data of considerable detail (e.g., task level or test item information) were often found to be recorded at Army Schools. These data, generally not forwarded to centralized files, are not routinely available for research purposes. However, at most of the schools it was possible to make arrangements for these raw data to be forwarded to the Project A data base (LRDB) manager. The format in which detailed training performance data is available varies by installation. It may be maintained in the TRADOC Educational Data System (TREDS) or on a local computer; it may be recorded in individual or class roster hard copy records; or it may be available only on the original individual test forms and score cards.

Development of Job Knowledge Tests

Job knowledge tests to be used as criterion measures of training performance in the 19 MOS are scheduled for development during the period October 1983 - December 1985. Development of these tests was begun during the visits to Army schools in which SME were interviewed, where the tasks to be represented in the new measures were identified as follows:

(1) Lists of approximately 300-600 tasks for each MOS were obtained from the Army Occupational Survey Program (AOSP). The lists provide the percentage of soldiers performing each task, by skill level.

(2) Tasks performed by 5 percent or fewer of the soldiers in Skill Level 1 were excluded from further consideration.

(3) Where subtasks or elements of tasks were listed separately in the AOSP, they were combined to generate whole tasks with a natural beginning and end (e.g., the elements "remove old tire" and "install new tire" were replaced by "change tire."

(4) Two hundred twenty-five tasks were selected by stratified random sampling. Duty categories (e.g., Redeye missile employment, Redeye operator maintenance, Redeye supply handling) were represented in proportion to the

number of tasks in each category in the total list. Task titles were put on cards, one per card, for sorting by SME.

(5) From three to six SME, depending on the number available at each Army installation, eliminated those tasks that were obsolete or unfamiliar to them. They sorted the remaining tasks on a 3-point scale of importance and a 5-point scale of frequency of performance errors.

(6) The 100 tasks with the highest combined importance/error ratings were selected for analysis in group discussions by SME and research personnel to generate statements of correct procedure and to identify the locus and characteristics of errors in performance. These descriptions of correct procedures and errors will be used to the extent possible to construct item stems, correct alternatives, and distractors for multiple-choice knowledge test items.

The ratings of tasks by SME for importance and error were analyzed to provide estimates of the consistency among raters in making the judgments. An appropriate form of intraclass correlation (ICC) where raters have not been selected randomly (Shrout & Fleiss, 1979) is:

$$\text{ICC} = \frac{\text{Mean Square Between} - \text{Residual Mean Square}}{\text{Mean Square Between} - (K-1)\ \text{Residual Mean Square}}$$

Reliability coefficients obtained using this formula are presented in Table 8 (Column A). Also included are the median Pearson r's for all paired comparisons of the raters (Column B) and an intraclass correlation based on a within-group design (Column C). The Column C coefficient was calculated because it is not affected by either small mean differences between tasks or lack of homogeneity of within-group variance. James, Wolf and Demaree (1981) have suggested that the intraclass correlation may underestimate interrater reliability in situations where there is little difference in mean ratings between targets (tasks) even though there is almost perfect agreement among ratings for each target. Interrater reliabilities for a single rater ranged from low to moderate. However, the reliability of most interest is the reliability of the average rating across SME. Using an average of four raters per MOS and applying the Spearman-Brown formula, the estimated reliabilities of the average ratings are shown in Column D.

The reason for the lack of high interrater reliability in some of the ratings of task importance and frequency of error in performance is not entirely clear. Rating distributions were frequently quite dissimilar across raters (e.g., many high ratings from one person and many low ratings from another in the same MOS), suggesting that the raters were using a different frame of reference, perhaps because of different prior experiences. Also, because job assignments vary after an individual leaves AIT, and because many SME lack current supervisory experience, precise judgments about importance and error rates may not always be possible.

Table 8

Task Rating Reliability Estimates

Post	MOS	Rating Type	A ICC	B Median γ	C $\bar{r}_{wg}$	D[a] Rel. of Av.
Ft. Dix	94B	Importance	.34	.42	.75	.74
		Error	.08	.08	.27	.26
Ft. Dix	63B	Importance	.16	.10	.56	.31
		Error	.10	.10	.33	.31
Ft. Dix	64C	Importance	.06	.09	.22	.29
		Error	.00	.02	.36	.09
Ft. Knox	19E	Importance	.07	.11	.68	.33
		Error	.12	.16	.50	.43
Ft. Bliss	16S	Importance	.14	.32	.57	.65
		Error	.15	.18	.55	.47
Ft. Gordon	05C	Importance	.24	.25	.55	.57
		Error	.18	.21	.66	.52

[a]Column D shows the reliability of the average rating over four raters if the correlation in Column B is taken as the average.

Summary and Conclusions

The training school site visits have produced a large fund of information to be used in the development of training achievement criterion measures. For those schools visited, we now have in hand detailed information on the current criterion measures, the way in which they are used, the procedures used or not used to store training school information, the objectives of the school, and the content and design of the curriculum. The tests currently being used are being systematically examined to determine how thoroughly they reflect training objectives and content and how useful they will be as sources of item content for the comprehensive knowledge tests that must be developed as part of this project.

The existing training measures are one major source of item content for the comprehensive knowledge tests. Another major source is the description of relevant job tasks that has been developed. We now have in hand a list of 200-250 job tasks per MOS that were sampled proportionately from the categories of tasks contained in the Army's occupational survey item bank and which have been refined in terms of their importance and relevance to the MOS in question.

Because the task descriptions were taken from the occupational survey item bank and the current Soldier's Manual, they are well anchored in the Army's design of the training curricula and the design of the job as it should be performed when the individual is required to perform in his or her specialty. While the importance and error ratings of the SME cannot be used to make precise discriminations among items, they are useful for identifying those tasks which are not currently job relevant and those which may be particularly prone to error.

As a consequence of the past year's effort, we are now in a reasonable position to begin generating the item pool for the comprehensive knowledge tests.

Next Steps

During the next 6-to-12 months the following activities will be paramount:

(1) The training school site visits will be completed, as per the original schedule in the Master Plan.

(2) The training objectives and training content will be matched (by the research staff) with the existing end-of-course-tests (EOCT) and with the task descriptions to determine where new items must be written.

(3) The comprehensive knowledge test item pool for each item will be generated.

(4) The items in each item pool will be submitted to SME and research staff review for a first determination of clarity, difficulty level, and relevance to specific training objectives.

(5) The items will be pretested with small samples of incumbents.

(6) The edited item pool will be administered to the criterion samples of trained vs. untrained enlisted personnel.

Abstract

As noted in the Introduction, an abstract of a related research report follows:

PRELIMINARY THOUGHTS ABOUT RESEARCH ON RELIABILITY AND VALIDITY OF ARMY TRAINING MEASURES*

R. L. Oxford-Carpenter and L. J. Schultz
(ARI)

This paper discusses the need to improve the reliability and validity of Army training measures and suggests essential steps in the improvement process. The Army problem that initiates the discussion is multifaceted. First, many Army trainers and test developers know little or nothing about assessment of test quality in general. Second, the state of the art is advancing so rapidly that there are constants in the quality of criterion-referenced tests, the Army's most prevalent type of training measures. Yet the Army--largely due to lack of psychometrically trained personnel--does not take advantage of these new developments. Third, little information on test quality is now being gathered on training measures. Fourth, many of these measures are psychometrically inadequate. Fifth, due to their inadequacy, such measures might detrimentally affect certain soldiers. Sixth, the overall picture of measurement on the training base is poorly documented--a problem that Project A is attempting to rectify. Seventh, the Army needs improved, multi-purpose training measures. For example, many training measures may not be useful for ranking purposes or for predicting future job performance. Methods for achieving these objectives are detailed in the paper.

*To be published as Selection and Classification Technical Area Working Paper 83-7.

VIII. CONSTRUCTION OF MOS-SPECIFIC CRITERION MEASURES

As noted in the Research Plan for Project A (ARI Research Report 1332), job selection research, in general, and military research, in particular, have been frequently criticized for the lack of job relevant validation criteria. Consequently, there has been considerable pressure to include such criterion measures in a comprehensive validation effort. Inclusion of such measures is also dictated by our model of overall soldier effectiveness. That is, successful execution of the specific job tasks for which an individual was trained is a significant component of overall effectiveness. It is necessary that we make every effort to assess this component of effectiveness as well as the state of the art will allow.

At the same time, we must also recognize that standardized hands-on task performance measurement is expensive and the R&D costs for developing such measures are also high. Consequently, trade-offs must be made. We have opted to devote considerable research effort to a smaller subset of MOS (i.e., nine) rather than compromise the amount of resources devoted to criterion development in each MOS beyond the point where the crucial research questions could be answered. The general strategy also includes the development of behaviorally anchored rating scales and paper-and-pencil knowledge test measures to determine if the less expensive methods can serve as substitutes for the more expensive.

Specific Objectives

The specific objective of this activity is to develop reliable, valid, and relatively economical measures of first- and second-tour job task performance of enlisted personnel in a sample of nine MOS. These measures will serve both as:

(1) Data collection instruments for establishing the relationships among various kinds of predictors and criterion measures, and

(2) Prototypes for the development of performance measures for additional MOS and/or MOS clusters.

Two different kinds of performance measures will be developed. The first will be direct measures of task performance (e.g., the average time it takes a soldier to troubleshoot and repair a malfunctioning electrical component). For measures of this kind, the incumbents must be evaluated under carefully structured and standardized conditions. The second kind will consist of two measures that are based on indirect evidence of performance: knowledge tests and ratings by supervisors or peers.

First Year Activities

During the first year our efforts relative to constructing MOS-specific criterion measures were focused principally on: (a) developing the specific samples of job tasks from which to build hands-on performance measures and

job specific knowledge tests; (b) using the critical incident method to develop rating scale measures of MOS-specific task performance; and, (c) using the accumulated task descriptions to develop a taxonomy of MOS-specific task performance categories, or factors, that will guide predictor selection and subsequent MOS clustering analyses.

Development of Task Samples for Performance Measurement

We began by generating a task sample for each of the four MOS in Batch A (13B, 64C, 71L, and 95B) by selecting and consolidating task statements from the Soldier's Manual (SM) and CODAP survey task descriptions.

These two job analysis sources were consolidated through a four-step procedure:

(1) Identify CODAP activities performed at SL1,

(2) Group CODAP statements under SM tasks,

(3) Group CODAP-only statements,

(4) Conduct Subject Matter Expert (SME) review.

Identify CODAP Activities Performed at Skill Level 1

The assumption for this step was that every activity included in the occupational survey questionnaire that had a nonzero response frequency, after allowing for error in the survey, was performed at skill level 1. The procedure for estimating the error was to compute the average response frequency for the survey and use that proportion to determine the boundaries of a confidence interval about zero. Activities with frequencies above the confidence interval were considered to have nonzero frequencies. For example, the confidence interval for 13B SL1 was $\pm$ 2.7. All statements with frequencies of 2.7 percent or lower were considered to be zero and were deleted from consideration; statements above 2.7 percent were considered part of the SL1 task domain. The results of this initial screen are shown in Table 9.

Group CODAP Statements Under SM Tasks

A CODAP statement (i.e., an item in the survey questionnaire) was placed under an SM (Soldier's Manual) task if the statement duplicated the SM task or was subsumed under the SM task as a step or variation in conditions. The effort first tried to identify SL1 tasks (either MOS specific or Common) with which the CODAP statement could be matched. If this could not be done, higher skill levels (HSL)--SL 2, 3, and 4--were successively reviewed and the CODAP statements matched with those SM tasks, if possible. Thus the grouping concentrated on matching CODAP statements with doctrine statements (i.e., Soldier's Manual tasks) wherever possible even if doctrine did not specifically identify the activity as a SL1 responsibility. All SL1 SM tasks were included regardless of whether or not they

Table 9

CODAP Statements Deleted From Task Domain

	13B	64C	71L	95B
CODAP Statements	669	677	822	776
Delete by "Zero" Frequency	67	169	329	210
(Confidence Interval)	(2.7)	(3.0)	(4.0)	(4.2)
Delete SME Review:				
Change in Doctrine	19		58	177
Nontask				20
Collective task	24			
Balance	559	508	435	369

had parallel CODAP statements. The number of SM tasks with CODAP statements and the number of CODAP statements that matched the SM tasks are shown for each of the four MOS by skill level in Table 10.

Group CODAP-Only Statements

Since some CODAP statements could not be matched with any SM task, or any subset of elements from an SM task, the third step was to edit the remaining CODAP statements so that although they were similar in format to the SM task statements, they were still a clear portrayal of additional task content not contained in the SM. In some cases a CODAP statement became a task statement by itself. In other cases a new task statement was developed which could appropriately subsume several CODAP statements. The results of this step are shown in Table 10.

Conduct SME Review

The final step in the consolidation was to confirm the grouping of CODAP statements with SME at the proponent school. At least three senior NCO or officers reviewed the grouping for each MOS. The review focussed on the placement of each CODAP statement and the appropriateness of the task titles for the CODAP-only tasks.

Some CODAP statements were deleted from the domain based on the SME review. As shown in Table 9, three reasons accounted for the deletions. The review of 13B identified changes in the doctrine (content specification) for the MOS that had occurred since the CODAP survey had been administered that would account for some of the CODAP-only tasks. Tasks that no longer applied (such as "Conduct ESC inspection") were deleted. The review of 95B identified administrative labels (such as "Question missing") that had been misconstrued as tasks. Also, in 13B, some ARTEP (collective) tasks were included. If the SME concluded that the collective tasks contained only individual tasks that were already in the domain, the statements were deleted. An example is "Fire high angle mission."

The result of the consolidation of SM and CODAP was a task domain for Skill Level 1 of each of the four MOS. The domain included:

(1) All SL1 tasks from the MOS SM and the SM of Common Tasks and their supporting CODAP statements.

(2) All HSL tasks with supporting CODAP statements

(3) All CODAP-only tasks.

These domains constitute a product in themselves in that they portray in precise task descriptive terms a definition of the job-world that an SL1 incumbent will face.

Table 10

Tasks in Performance Domain

	13B	64C	71L	95B
SM MOS TASKS				
SL1 w/CODAP (#)	67 (91)	21 (220)	39 (130)	98 (175)
SL1 w/out CODAP	55	1	0	40
SL2 w/CODAP (#)	24 (15)	*	*	15 (27)
SL3 w/CODAP (#)	28 (39)	2 (3)	50 (93)	0 (0)
SL4 w/CODAP (#)	19 (21)	3 (5)	1 (1)	2 (3)
COMMON TASKS				
SL1 w/CODAP (#)	22 (42)	61 (69)	61 (61)	46 (49)
SL1 w/out CODAP	48 **	10	10	23
SL2 w/CODAP (#)	13 (34)	20 (29)	10 (10)	5 (5)
SL3 w/CODAP (#)	5 (10)	5 (5)	2 (2)	3 (4)
SL4 w/CODAP (#)	2 (4)	10 (12)	1 (1)	2 (2)
CODAP ONLY (#)	73 (303)	33 (165)	29 (137)	70 (104)
TOTAL DOMAIN	356 (559)	166 (508)	203 (435)	304 (369)

* MOS combines SL1 and SL2. **One common task in MOS SM.

Note: Numbers in parentheses are CODAP statements subsumed by SM tasks.

Table 11

Effects of Narrowing Domain

	13B	64C	71L	95B
TASKS IN DOMAIN (CODAP)	356 (559)	166 (508)	203 (435)	304 (369)
COMBINE SYSTEMS		NA	NA	NA
SL1	37			
SL2	10			
SL3	10			
SL4	5			
RESTRICTED DUTY POSITION		NA		NA
SL1 (CODAP)	46 (8)		19 (66)	
SL2	0		--	
SL3	0		14 (18)	
SL4	0		1 (1)	
CODAP-only	0		8 (18)	
DESIGNATED NA FOR SL1		NA	NA	NA
SL2 (CODAP)	0			
SL3	5 (6)			
SL4	13 (20)			
LOW FREQUENCY				
SM MOS Tasks			NA	NA
SL2 (CODAP)	1 (1)	--		
SL3	3 (4)	2 (3)		
SL4	1 (14)	3 (5)		
Common Tasks				
SL2 (CODAP)	9 (29)	13 (14)		
SL3	5 (10)	4 (4)		
SL4	1 (1)	9 (9)		
CODAP-only	34 (105)	16 (83)		
PRELIMINARY SORT	NA	NA	NA	
SL1				93 (148)
SL2				18 (30)
SL3				2 (3)
SL4				4 (5)
CODAP-only				59 (193)
TASKS DELETED (CODAP)	180 (185)	47 (118)	42 (103)	176 (279)
TASKS FOR CRITICALITY	176 (374)	119 (390)	161 (332)	128 (90)

Narrow Domain

The task domains that were assembled were still too broad to assure that any task selected for hands-on test development would be both frequently performed and critical to job performance. The domains were, therefore, narrowed further through a six-step process. The goal was to arrive at a maximum number of tasks that could be managed feasibly in a systematic review by SME for criticality and clustering decisions. Because each of the four MOS presented unique structures, resources, and requirements, not all of the six steps that follow were performed for each MOS:

(1) Combine system specific tasks

(2) Delete tasks that pertain only to restricted duty positions

(3) Delete HSL tasks that have been officially designated not relevant to SL1 job performance by proponent

(4) Translate CODAP frequencies into task frequencies

(5) Delete HSL and CODAP-only tasks with atypically low frequencies

(6) Collect preliminary criticality ratings.

The tasks deleted as a result of each step are summarized in Table 11, as are the number of tasks in the final set selected for criticality evaluation.

Combine systems-specific tasks. The Soldier's Manual for 13B treated the same operations performed on different equipment systems as separate tasks. For example, "Measure the quadrant with the range quadrant" applies to howitzers and is treated as six tasks. From a training perspective that is appropriate because the performance steps vary somewhat among the howitzers. From the perspective of this project, however, treating such tasks as one rather than six tasks was preferable. The justification was that a soldier could only be held accountable for performing the task on the one kind of howitzer in his unit. If the task "Measure the quadrant with the range quadrant," should be selected, project staff may have to prepare as many as six forms of the test, but it should represent only one task in the criterion space.

Delete for restricted duty positions. The criterion for deleting a duty position task was that an Additional Skill Ident fier or Special Skill Identifier and at least one week of special training were specified as being required for task performance. Only the 13B and 71L domains included duty positions that met that criterion. There were four duty positions for 13B: Artillery Mechanic (M198), Assembler; 155mm Atomic Projectile, Assembler; 8-Inch Atomic Projectile; and Nuclear Security Guard. The only 71L duty position that met the criterion was postal clerk.

Delete Higher Skill Level (HSL) tasks designated not applicable to Skill Level 1 (SL1). A set of MOS tasks for 13B had been reviewed by an

Table 11

Effects of Narrowing Domain

	13B	64C	71L	95B
TASKS IN DOMAIN (CODAP)	356 (559)	166 (508)	203 (435)	304 (369)
COMBINE SYSTEMS		NA	NA	NA
SL1	37			
SL2	10			
SL3	10			
SL4	5			
RESTRICTED DUTY POSITION		NA		NA
SL1 (CODAP)	46 (8)		19 (66)	
SL2	0		--	
SL3	0		14 (18)	
SL4	0		1 (1)	
CODAP-only	0		8 (18)	
DESIGNATED NA FOR SL1		NA	NA	NA
SL2 (CODAP)	0			
SL3	5 (6)			
SL4	13 (20)			
LOW FREQUENCY				
SM MOS Tasks			NA	NA
SL2 (CODAP)	1 (1)	--		
SL3	3 (4)	2 (3)		
SL4	1 (14)	3 (5)		
Common Tasks				
SL2 (CODAP)	9 (29)	13 (14)		
SL3	5 (10)	4 (4)		
SL4	1 (1)	9 (9)		
CODAP-only	34 (105)	16 (83)		
PRELIMINARY SORT	NA	NA	NA	
SL1				93 (148)
SL2				18 (30)
SL3				2 (3)
SL4				4 (5)
CODAP-only				59 (193)
TASKS DELETED (CODAP)	180 (185)	47 (118)	42 (103)	176 (279)
TASKS FOR CRITICALITY	176 (374)	119 (390)	161 (332)	128 (90)

Artillery Center Critical Task Board just before the SME review of the task domain. The results of that Board were distinctive in that the Board assigned levels of performance to each task by skill level rather than assuming a clear break between skill levels. Eighteen HSL tasks that had been in the task domain for SL1 had been rated in the lowest category (not applicable for Skill Level 1). Those tasks were deleted. Ratings for three tasks (after combining for weapons systems) which had not been in the domain (because no CODAP data covered them) had ratings that indicated that Skill Level 1 soldiers should at least have some knowledge of the task. Those tasks were added to the domain and are included in the domain totals in Table 10.

Translate CODAP frequencies into task frequencies. CODAP statements did not always correspond directly with task statements. In some cases, the CODAP statements represented steps within the tasks. In other cases, the CODAP statements represented various conditions. For example, CODAP frequencies covered statements like "Drive vehicle 2 1/2 tons or less in administrative convoy" and "Drive tractor-trailer combination vehicle in tactical convoy" when the 64C MOS task was "Operate Vehicle in Convoy." In still other cases, the CODAP statement was equipment specific while the task for testing purposes was generic. For example, CODAP frequencies covered "Prepare semifixed ammunition" and "Prepare separate loaded ammunition," but the task for consideration was "Prepare ammunition."

The algorithm for assigning frequencies to tasks is shown in Figure 6. Generally, when CODAP and task statements matched, the frequency for the matching statement was applied to the task. If there was no match, the most frequent step or condition was the basis for the task frequency. However, in some cases, frequencies were aggregated to account for equipment differences.

Delete low frequency HSL and CODAP-only tasks. The purpose of this screen was to identify tasks with atypically low frequencies. The general approach was to compare frequency distributions of the Skill Level 1 tasks (MOS and Common) with the HSL and CODAP-only tasks. HSL and CODAP-only tasks were then eliminated until the two groups were not significantly different with respect to location, dispersion, and form.

A four-step procedure identified the atypically infrequent tasks to be elminated:

(1) List the response frequencies of Skill Level 1 tasks.

(2) List the response frequencies of HSL/CODAP-only tasks.

(3) Test groups for difference using Mann-Whitney U test.

(4) If groups were different and the HSL/CODAP-only group had tasks with lower response frequencies, eliminate lowest frequency tasks until group differences were not significant at .01 level.

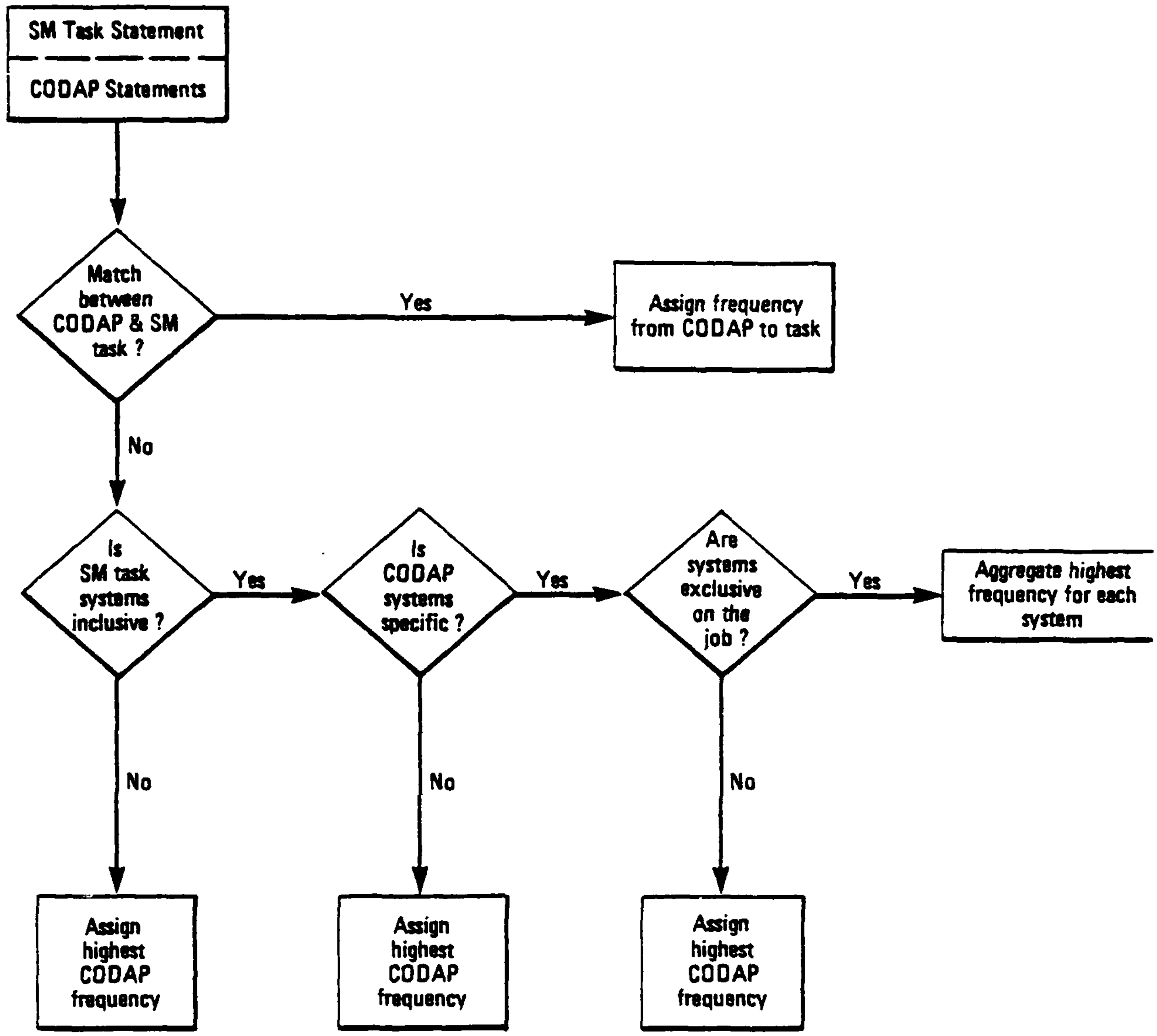


Figure 6. **Method for Assigning Frequencies to Tasks**

Collect Preliminary Criticality Ratings

Because the 95B Skill Level 1 domain was so large, it was narrowed through a preliminary sort on criticality. Ten senior 95B NCO were given 304 cards with task titles and brief descriptions of the scope of each task. They were asked to sort the tasks into two groups of approximately the same size. One group contained the more critical tasks, the other the less critical. They then ranked the group of more critical tasks from most to least important within that group. The ratings assigned to each task by the NCO were combined and plotted against respective CODAP frequencies to select the most critical, most frequently performed task for 95B SL1.

Scale Criticality and Cluster Remaining Tasks

The tasks remaining for each MOS after narrowing the domain were the candidates for selection. Since only a subset of the candidates could be covered in the large-scale data collection, further information was gathered to enable selecting a subset that contained the most critical tasks and represented the functional areas of the narrowed domain.

This information was gathered through a two-stage data collection effort with 15 senior NCO and officer SME at each of the four proponents. During the first phase the SME ranked each task. Each SME was given a card for each task. The card had the task title and a brief description of the scope of the task. The SME selected the one task that was most important for a European combat defensive situation and the one task that was least important for that situation. The SME repeated the process until all tasks were ranked.

During the second phase the SME sorted the tasks into groups based on the performance requirements of the tasks. The SME worked with the same cards as in the criticality phase. The results for the 15 raters were analyzed by means of a hierarchical clustering program.

For the final task selection project staff selected tasks to represent the clusters, giving priority to high criticality/high frequency tasks.

Summary and Next Steps

What we have produced at this point is a very carefully specified set of tasks that will form the content of the hands-on performance measures and the paper-pencil job knowledge measures for the Batch A MOS. Great care was taken with the above procedural steps to insure that the content of the two MOS-specific performance measures accurately reflects relevant job content that is highly representative of what people actually do and that is critical for effective performance in the MOS.

In the coming year we will repeat the process for Batch B MOS and will begin to develop the actual exercises that will test for proficiency on the specified tasks.

MOS-Specific Behaviorally Anchored Performance Rating Scales

As noted earlier, two alternative methods will be used to assess MOS-specific job performance. One method will use paper-and-pencil measures of job knowledge that reflect the tasks being assessed by the hands-on measures. Construction of these tests will begin during FY84.

The second method uses the critical incident technique to develop behaviorally anchored rating scales for task performance factors in specific MOS. Work began on these measures during the last quarter of FY83 and is continuing at the present time.

Procedure

To develop behaviorally anchored rating scales for the four MOS in Batch A (13B, 64C, 71L, 95B), critical incident workshops have been conducted with 8-15 NCO from each MOS in each of four locations. From these incidents an initial set of performance factors for each MOS has been constructed by having the project staff group critical incidents into categories that are judged to represent similar elements of task performance. The retranslation phase of the procedure and the completion of the rating scales will be done in FY84.

Next Steps

During the next contract period the remaining workshops will be conducted and the development of specific performance factors for each MOS in the Batch A and Batch B samples will be completed. The remaining steps will include the all-important retranslation step (see Research Plan), which helps to insure valid and reliable performance dimensions. The judgments obtained in the retranslation step will be used to construct rating scales for each of the MOS-specific performance dimensions. These new rating scales will then be pilot tested on small samples of incumbents. The result will be the first set of behaviorally defined rating scales that has ever been used to assure specific technical performance in a skilled job.

Abstract

As noted in the Introduction, an abstract of a related research report follows.

ISSUES AND STRATEGIES IN MEASURING PERFORMANCE IN ARMY JOBS*
W. C. Osborn
(HumRRO)

Job-specific performance, the measure of soldier readiness of primary concern to the Army, is the criterion behavior of central interest to the research project. The objective in this regard is to develop reliable, valid, and economical measures of first- and second-tour job performance of soldiers in the sample of Army jobs.

Methods of performance measurement will be examined in the early stages of the research in an effort to find mixes tailored in measurement efficiency to the performance requirements of various Army jobs. Methods of measurement vary in efficiency--validity and feasibility combined--depending on the kind of behavior being measured. Can do or maximum behavior is best measured by test, but methods of testing differ in efficiency. Performance testing is the only valid measure of proficiency for some job tasks, thus it justifies the added cost of the method. For other tasks, job knowledge tests provide adequate validity while offering substantial savings in testing time and resources. Will do or typical behavior, on the other hand, can be tested but only by unobtrusive means--means which are costly at best and unethical at worst. Dimensions of typical job behavior are, therefore, normally measured indirectly by supervisor or peer ratings.

Presented in this paper are the procedures by which test and rating methods are judged suitable to various elements or aspects of job performance. Also described are task analytic and test or rating development techniques necessary to assure a valid bridge between job content and performance measure. Plans are presented for empirically evaluating relationships among measurement methods and types of behavior, data that will then enable the selection of an efficient and comprehensive mix of measures for final criterion measurement.

---

*Paper presented at the 1983 Annual Convention of the American Psychological Association in Anaheim, California, August 1983.

IX. DEVELOPMENT OF ARMY-WIDE JOB PERFORMANCE CRITERIA

This part of the effort is devoted to the identification, refinement, and development of Army-wide performance measures. Army-wide performance measures are those indicators of general performance and effectiveness not related directly to the performance of MOS-specific tasks.

The central goals of this activity are: (a) to identify aspects of soldier effectiveness that apply to all MOS; (b) to identify and/or develop valid indicators to measure these aspects of effectiveness; and (c) to establish the indicators as criteria of soldier effectiveness and, where appropriate, as in-service predictors of future performance or other aspects of soldier effectiveness. In-service predictors are measures obtained after a soldier enters the Army; they predict the soldier's later performance or effectiveness in his/her military career. Measures must be identified and/or developed for both first-tour and second-tour performance.

Definition of Army-wide effectivenss within the general overall model of soldier effectiveness requires careful specification of the relevant criterion space. "Outcome indicators" and objective administrative indexes such as attrition, disciplinary actions, special awards, schools attended, etc., are clearly Army-wide criteria, and measures of these types of criteria are of concern in the research. A second focal point is the development of general performance and soldier effectiveness measures. An individual's "Worth to the Army" is conceptualized as including a relatively broad set of soldier effectiveness criteria such as organizational commitment, organizational socialization, and morale.

Special behavior-based rating scales are being prepared to measure soldier effectiveness on all important dimensions identified in the initial model development work, and supervisory, peer, and self ratings will be gathered to provide a set of Army-wide effectiveness criteria.

The Preliminary Model

To generate the initial model for the general effectiveness domain we made some preliminary hypotheses about constructs that might be considered. These constructs focus on the areas of organizational commitment, organizational socialization, and morale.

Organizational Commitment--The concept of organizational commitment (Porter, Steers, Mowday, & Boulian, 1974; Steers, 1977) refers to the strength of a person's identification with and involvement in the organization. It incorporates three kinds of attitudinal and cognitive elements: acceptance and internalization of organizational values and goals, motivation to exert effort toward the accomplishment of organizational objectives, and firm intentions of staying in the organization. It connotes a sense of loyalty to the organization as a whole and a desire to fulfill more general role requirements that come with organizational membership.

Organizational Socialization--Organizational socialization is the process by which an individual acquires the social knowledge and skills necessary to

assume an organizational role (Van Mannen & Schein, 1979). Some part of this knowledge and skill is, of course, job specific. For example, training programs designed to improve the effectiveness with which a person performs job-realted tasks are part of the process of organizational socialization. But there are also many other knowledges and skills necessary for effective functioning as an organizational member that are not job specific. When the socialization process is successful, a person will acquire not only job-related skills but also new patterns of behavior with subordinates, peers, and superiors in the organization; new attitudes, beliefs, and values in line with organizational norms; and new ways of using time not formally dedicated to performing job-related tasks.

Morale--The concept of morale has traditionally been regarded as an extremely important element in military organizations. The concept of military morale is multifaceted. It seems to involve feelings of determination to overcome obstacles, confidence about the likelihood of success, exaltation of ideals, optimism even in the face of severe adversity, courage, discipline, and group cohesiveness. (Motowidlo, Dowell, Hopp, Borman, Johnson, & Dunnette, 1976).

Our preliminary hypotheses then, were that soldiers who show high levels of commitment to the Army, acceptance of Army norms, and morale are more effective soldiers in this broader sense and are also of more value to the Army. Sixteen dimensions were identified within these three constructs. They are listed in Figure 7. Also listed in Figure 7 are the 22 dimensions derived empirically. A description of that procedure follows.

Development of General Effectiveness Measures

The principal means being used to build new measures of general soldier effectiveness is the behavioral analysis or behaviorally anchored rating scale (BARS) technique. It is dependent on the gathering of critical incident descriptions of job behaviors. It is also the principal means by which the model or theory of general effectiveness will be refined, revised, and developed.

At this time, we have conducted two BARS development workshops with a total of 14 experienced Army officers (captains and majors). In the workshops these officers generated 245 examples of first-tour soldier effectivenes, and we have performed a preliminary content analysis to explore possible dimensions emerging to define soldier effectiveness. Several other workshops will be conducted with officers and NCO to ensure good coverage of the entire target domain, but these 245 examples provide some idea of what that domain will look like.

Twenty-two relatively fine-grained and specific dimensions were derived from the content analysis. They appear listed in Figure 7.

Comparison of Model Dimensions and the Empirically Derived Dimensions

To obtain an initial idea about how the 22 empirically derived dimensions might fit into the dimensions identified in the initial model of soldier effectiveness, we sorted each behavioral example into one of the model's

Preliminary Dimensions

1. Following orders
2. Following regulations
3. Respect authority
4. Military bearing
5. Commitment
6. Cooperation
7. Comradery
8. Concern with unit goals
9. Boosting unit morale
10. Leadership
11. Perserverance
12. Endurance
13. Conscientiousness
14. Initiative
15. Discipline
16. Other

Empirically Derived Dimensions

1. Promptness vs. Tardiness
2. Job Knowledge
3. Personal Financial Management
4. Stealing, Lying, Sociopathy
5. Physical Fitness
6. Maintaining Clean and Neat Quarters/Environment
7. Drug/Alcohol Abuse
8. Maintaining Own Equipment
9. Attention to Detail
10. Following Standard Operating Procedures on Tasks
11. Initiative/Volunteering
12. Perserverance
13. Effort to Improve Soldiering and Job Skills
14. Military Appearance
15. Accepting Orders from Superiors
16. Military Courtesy
17. Following Regulations
18. Leadership: Taking Initiative to Lead Others; Taking Charge When Placed in Leadership Position
19. Leadership: Motivating Others to Push On vs. Encouraging Them to Goof Off
20. Leaderhip: Correcting Performance of Others
21. Leadership: Instructing Others
22. Displaying Concern for Individual Others and the Unit

Figure 7. Preliminary dimensions and empirically derived dimensions of soldier effectiveness

dimensions. Then, behavioral incident membership in dimensions within the two systems was cross-referenced to provide a rough comparison between the two dimensional systems. This comparison is shown in Figure 8.

Results of this cross-referencing show first that two dimensions in the model of soldier effectiveness are not reflected in any examples. Commitment and Comradery had no incidents sorted into them. Second, Conscientiousness and Following Regulations are probably too broad, with incidents from 7 of the 22 empirical dimensions appearing in each of these categories of the model.

Third, four of the empirical dimensions do not have any representation in the model. Job Knowledge/Skill, Financial Management, Stealing/Lying, and Physical Fitness are not reflected in the model's dimensions. Fourth, there are some near one-to-one matches between the two dimensional systems. Military Bearing and Military Appearance; Boosting Unit Morale and the second Leadership dimension in the empirically derived dimensions, Perseverance (in both systems); and Discipline and Drug/Alcohol Abuse provide good matches. In the last case, however, Discipline is defined much more broadly than the content represented in the Drug Abuse examples.

Finally, and this is perhaps the most salient result, there seems to be considerable overlap between the dimension content in the two systems, but often the configuration of that content differs. Esentially, elements of the dimensions in the two systems are put together differently.

The most important objective in developing dimensions is to achieve the purposes of the project, and this overriding concern will guide future efforts to integrate empirical information with the theoretical model. Dimensions will be developed and defined to reflect in a comprehensive and, at the same time, efficient manner, the domain of soldier effectiveness. Dimensions will be derived to provide raters using rating scales based on the dimensions with an easy-to-understand, highly face-valid rating format that reflects accurately the behavioral requirements of this domain.

This approach has the advantage of forcing a broad perspective on the criterion domain. It points out potentially important elements of individual effectiveness that might be overlooked by purely inductive approaches to job and task analysis. For this reason, we believe the model is useful for guiding efforts to impose structure upon the complexity of what "soldier effectiveness" might mean in the Army.

Development of Archival Records as Army-Wide Criterion Measures

A major activity within our overall program of performance criterion development is to explore the use of archival administrative records in the formation of first-tour criteria and in-service predictors of soldier effectiveness. The Enlisted Master File (EMF), the Offical Military Personnel File (OMPF), and the Military Personnel Records Jacket (MPRJ) are the records sources that contain administrative actions that could be used to form measures of first-tour soldier effectiveness.

Figure 8

Comparison of the Model's Dimensions and the Empirically Derived Dimensions

Model Dimensions \ Empirical Dimensions	1. Promptness	2. Job Skill	3. Finances	4. Stealing/Lying	5. Physical Fitness	6. Clean Quarters	7. Drug Abuse	8. Equipment	9. Detail	10. Following SOP	11. Initiative	12. Perseverence	13. Improve Skills	14. Appearance	15. Accept Orders	16. Courtesy	17. Follow Regs.	18. Leadership I	19. Leadership II	20. Leadership III	21. Leadership IV	22. Concern for Others
1. Following Orders						0			0						■							
2. Following Regulations	■							X	0					0			■	X		X		
3. Respect Authority														0		■	X					
4. Military Bearing														■								
5. Commitment																						
6. Cooperation											X										■	■
7. Comradery																						
8. Concern Unit Goals		0							0													0
9. Boosting Unit Morale																			■			0
10. Leadership										■								X				
11. Perseverence											0	■						0				
12. Endurance								0			0	0										
13. Conscientiousness						■		X	■		0	0	■					X				
14. Initiative		0									X							X		0		
15. Discipline							■															
16. Other		■	■	■	■						0									0		

■ = Very high overlap (>75% of behavioral examples falling in empirical dimension fall in model dimension)
X = Intermediate amount of overlap (15-74% of examples falling in empirical dimensions fall in model dimensions)
0 - Small overlap (at least 1 example but <15% of examples falling in empirical dimensions fall in model dimensions)
Blank - No overlap.

As mentioned in ARI Research Report 1332, a serious difficulty in using administrative records to form soldier effectiveness criteria is that the material in the records very often reflects only exceptionally good or exceptionally poor performance. Measures of performance based on infrequently appearing personnel actions could have very little variance, i.e., almost everyone has the same score. A strategy for dealing with the skewness in records data that results from low base rates is to combine records of different kinds of events and actions into more general indexes. When scores on administrative measures that reflect the same underlying constructs are combined, the base rate might improve to a level where significantly higher correlations with other variables would be possible. Consequently, before administrative records composites can be formed and assigned to performance constructs we must determine which administrative indexes have sufficient variance and acceptable base rates to warrant inclusion in composite formation, and which records distinguish effective from ineffective soldier performance. As such, we must identify which administrative actions reflect Army-wide soldier effectiveness and from which archival sources it is most feasible to obtain them.

Accordingly, during the past six months we have begun a detailed examination of the three archival data sources and an analysis of the feasibility of developing criterion indices from them. The Enlisted Master File is a computer file corresponding to every enlisted individual currently on the U.S. Army payroll. It contains a large number of variables for each individual ranging from pay grade to Skills Qualification Test (SQT) scores to appraisal ratings in the form of the Enlisted Efficiency Report (EER). A complete description of the variables available from the EMF is given in the Longitudinal Research Data Base (LRDB) plan.

An initial review of the EMF was carried out by interviewing several key Army personnel who have knowledge of and/or responsibility for the EMF. The variables which appeared to hold the most promise are: (1) reason for separation, (2) reenlistment eligibility, (3) reenlistment eligibility bar, and (4) weighted Enlistment Evaluation Report score. With the exception of the weighted EER, these measures may more appropriately be considered outcomes that result from performance, rather than evaluations of performance per se. In theory, the EER variable on the EMF, which is a weighted average of a soldier's last five EER should be an excellent variable. As a practical matter, however, its usefulness may be limited. Since EER are only done on soldiers in grades E5 and above, only a small percentage of the first-tour cohort is likely to have had even one EER at the time of the data collection. Secondly, in the past few years EER scores have tended to cluster at the maximum of 125. Thus, distinguishing effective from ineffective performers on the basis of EER scores may not be possible. A definitive answer regarding suitability of EMF variables for use as criteria is dependent on our own comprehensive examination and analysis of the existing computer records and existing EMF documentation. That analysis is currently in progress.

Information in the Official Military Personnel File (OMPF) is stored on microfiche. Depending upon their purpose, documents are filed in one of three sections:

(1) The performance (P) fiche. The P fiche is the portion of the OMPF where performance, commendatory, and disciplinary data are filed.

(2) The service (S) fiche. The S fiche is the OMPF section where general information and service data are filed.

(3) The restricted (R) fiche. The R fiche is the OMPF section for historical data that may be biased against the soldier when viewed by selection boards or career managers. For this reason release of information on this fiche is controlled.

The initial examination of microfiche records was conducted by project staff. They conducted a three day site visit at the Enlisted Records and Evaluation Center (EREC) at Ft. Benjamin Harrison. A total sample of 465 individual soldiers was drawn from a variety of MOS. If a microfiche packet could be found for the individual, each record in the packet was examined by a staff member and a variety of information items were recorded. A summary of the major findings is as follows:

(1) Of the 414 microfiche packets that could be located, 278 contained only a serivce fiche while 136 contained both a service and performance fiche.

(2) Of the 136 soldiers in our sample who had performance fiche, 44 of them (32 percent) were prior service members. Of these 44 soldiers, 20 had EER in their files. Six of the soldiers had two EER apiece for a total of 26 EER. The distribution of EER scores was:

Frequency	Score
13	125
3	123-124.9
5	121-122.9
5	121

(3) A total of 52 Articles 15 is issued to the 136 soldiers who had a performance fiche.

(4) 63 awards were received by the 136 soldiers. 41 of these awards were for completion of a training course.

(5) 12 letters of appreciation/commendation appeared on the performance fiche.

(6) Of the 136 soldiers, 26 were credited with having attended a school. Two of these soldiers attended two schools apiece.

After examining the microfiche and the regulations governing their composition as well as interviewing cognizant officials, we reached two conclusions:

(1) The data which exist in the OMPF are not nearly as complete or timely as we would like them to be. For grades E5 and below, which are the grade levels that enlisted personnel will be in the FY83/84 first-tour cohort study, there is an 8-12 month backlog from the time a personnel action is taken to the time it appears on microfiche at EREC.

(2) Whether performance-related material for a given soldier appears in the OMPF depends in large part on his or her CO. If a commendatory or disciplinary action is taken on a soldier, the CO has three choices. He/she can either send it to EREC to be filmed on the soldier's performance fiche, his restricted fiche, or neither. We did not see the restricted fiche and, given their sensitive nature, it is questionable at this time whether we will gain access to these fiche. Keeping in mind the 8-12 month backlog, even if we are granted permission to view the restricted fiche, administrative index data may not be available when we need them. The CO's third alternative is of greatest concern. While AR 640-10 lists specific disposition of each document authorized to appear in the OMPF, the individual CO has discretionary power regarding which commendatory letters, letters of reprimand, and Articles 15, for grades E5 and below, get forwarded to EREC for inclusion on the OMPF. It is therefore possible for a soldier not to have a performance fiche but have one or more Articles 15 in his Military Personnel Records Jacket (MPFJ).

Because of the limitations in the microfiche records, determination of the discrepancy in type, quantity, quality, and timeliness of the information contained in a soldier's MPFJ (201 file) and the information that exists in the OMPF appears to be of vital importance. The MPRJ (201) file is the primary mechanism for storing information about an individual's service record. It is the most complete and up-to-date record and it physically follows the individual wherever he or she goes. It is located at the Military Personnel Office (MILPO) that serves the soldier's unit.

Abstracts

As noted in the Introduction, abstracts of relevant and related research reports follow:

PUTTING THE "DOLLARS" INTO UTILITY ANALYSES*

N. K. Eaton, H. Wing, and K. J. Mitchell
(ARI)

Estimating the dollar benefit of performance improvement due to selection testing requires estimation of SD$, the standard deviation of performance measured in dollars. One method is to compute SD$ from estimates of the dollar value of performance at various levels. Unfortunately, such estimates are troublesome where management does not typically place dollar values on productivity or where complex, expensive equipment is used. Two alternate techniques were developed. One uses estimates of relative productivity for workers at various performance levels, but the dollar value of average performance only. The other focuses on changes in the numbers and costs of employee/machine units operating at given levels in a system and on the consequent impact on overall system costs. All three were applied to an example using the U.S. Army tank system. The results suggest the appropriateness of the two techniques developed in this research.

DEVELOPING A MODEL OF SOLDIER EFFECTIVENESS: A STRATEGY AND PRELIMINARY RESULTS*

W. C. Borman
(PDRI)
S. J. Motowidlo
(College of Business Administration, Pennsylvania State University)
L. M. Hanser
(ARI)

This paper introduces a model of individual effectiveness that extends beyond successful performance on specific job tasks and on directly job-related effectiveness dimensions. The model of soldier effectiveness suggested here contains elements of morale, along with organizational commitment and socialization. The notion is that these broad constructs represent important criterion behaviors that contribute to an individual soldier's "worth to the Army" and to his/her unit's organizational effectiveness. Fifteen dimensions springing from the model are named and defined.

The paper also presents preliminary results of behavioral analysis or BARS (Smith & Kendall, 1963) research to develop dimensions of soldier effectiveness using this comparatively inductive procedure. Fourteen Army officers in two workshops generated a total of 245 behavioral examples of soldier effectiveness in these early stages of the research project. Although by no means a formal test of the solder effectiveness model, the content of the examples generated showed considerable similarity to many elements of the model. Exceptions were noted and discussed. Also discussed were certain advantages to taking a broader perspective in studying individual effectiveness, particularly in this kind of organization, as well as risks inherent in considering criterion elements that are not directly job-related.

---

*Paper presented at the 1983 Annual Convention of the American Psychological Association in Anaheim, California, August 1983.

DUSTING OFF OLD DATA:
PROBLEMS ENCOUNTERED WITH ARCHIVAL RECORDS*

L. M. Hanser and F. C. Grafton
(ARI)

Frequently validity research must be completed using data which already exist, because the time or expense of developing new data is prohibitive. Under the circumstances, records of scores on the selection device of interest may be fairly complete. They may require some brushing off or perhaps some cleaning up, but usually a minimum amount of effort is involved. The criterion, on the other hand, is usually more troublesome. This paper reports on the problems encountered in validity research conducted by the U.S. Army Research Institute using archival data. The intent is not to report on the results of this research, but rather on the problems encountered as the result of using archival data, and how those problems were addressed.

The Army's current selection and classification instrument is the Armed Services Vocational Aptitude Battery (ASVAB). This battery currently contains ten subtests. These subtests are combined to provide a score on the Armed Forces Qualification Test (AFQT), which is used for selection, and to provide 10 Aptitude Area Composites, nine of which are used currently for classification. In a given year over 125,000 individuals enlist in the Army. These individuals are each placed in one of over 200 entry-level occupations.

In undertaking research to determine the validity of ASVAB, suitable criterion measures were sought. For several reasons, this presented a problem. Most individuals enlist for a three- or four-year period. During their first four or more months they are engaged in basic and advanced training. Most training is mastery learning by design and yields little in the way of criterion measures. Those measures collected during training do not typically make their way into long-term storage files. Upon completion of training, each individual is assigned to one of hundreds of duty stations spread around the world. This dispersal makes the collection of criterion data difficult at best.

Very little information which might be useful as a criterion is collected in centralized files. Even progression of promotions is difficult to track. A measure with possible utility as a criterion was found to exist in centralized records. This measure is the score on Skill Qualification Tests (SQT), which are routinely given approximately 16 months after entry into the Army. Problems encountered in actually using these scores as a criteron are addressed also.

---

*Paper presented at the 1983 Annual Convention of the American Psychological Association in Anaheim, California, August 1983.

X. PREDICTOR SELECTION AND DEVELOPMENT

General Purpose

The general purpose of this activity is to identify an efficient and effective set of initial or preinduction predictors of soldier performance. By efficient, we mean that time and money to be expended on operational administration of the predictors is kept as low as possible, and by effective, we mean that the predictors forecast as accurately as possible the degree of success to be expected of recruits in various aspects of soldier performance, e.g., overall adaptation to the Army, completion of training, performance in specific MOS, and reenlistment.

There are two different, but related aspects to this general purpose. First, we will evaluate the effectiveness of the present set of initial predictors used by the Army contrasted with a more comprehensive array of criteria than has been used previously. We will identify and develop new predictors that are most likely to be effective and efficient additions to the present set of predictors. The validity or effectiveness of these new predictors will be investigated in the same way as the validity of the present set of predictors. The evaluation of the efficiency of newly developed predictors will require analysis of the improvement in prediction of soldier performance gained by use of the new predictors over that obtained by the sole use of the present set of initial predictors.

A major activity this year was a comprehensive literature search and review. The search was conducted by three research teams, each responsible for a broadly defined area of human abilities or characteristics. The three areas were cognitive abilities; noncognitive charcteristics such as vocational interests, biographical data, and measures of temperament; and psychomotor/physical abilities. These areas or domains proved to be convenient for purposes of organizing and conducting literature search activities, but were not used as (nor intended to be) a final taxonomy of possible predictor measures.

The output of the literature search served as input for (a) the selection of the preliminary battery, (b) the writing of the literature review report, (c) the formulation of a comprehensive model of the predictor space in the form of specifying the predictor constructs that seem to best describe the latent variables measured by the available tests, and (d) the development of the formal technical review that will begin in October 1983 (FY84).

Considerable staff time was devoted to defining the total array of constructs that seemed to account for the total predictor space. In a very real sense this was an important step in "theory development" as it pertains to the measurement of individual differences of Army applicants. It is also the array of constructs that will be used by the expert judges in the technical review to scale the expected relationships between the predictor constructs and the array of criterion factors that our current model says constitute total performance space.

The array of criterion factors was produced by the MOS job analyses, the critical incident workshop, the review of archival records, and the analysis of the AIT programs of instruction.

Again, considering both predictor space and criterion space in construct terms has been extremely valuable in our development work so far, and will continue to be so as we refine and expand our knowledge about both these domains through the major phases of the project.

A second major activity conducted during the first year was the identification and development of the Preliminary Battery. This battery is intended to be an efficient, comprehensive set of predictors not covered by the present Army preinduction measures. Its administration to trainees will allow an empirical determination of the extent to which additional, conceptually distinct predictor measures actually measure different human abilities than are currently measured and, through follow-up research, the extent to which such measures add precision to the prediction of success in training performance and on-the-job performance.

The content of the Preliminary Battery was carefully chosen in as efficient a manner as possible to be as comprehensive as possible. The research staff first compiled a list of all even remotely appropriate measures identified in the literature search. This was called "List 1"; it was screened by eliminating measures according to several "knockout" factors. That is, the following factors were used to eliminate potential predictors from further consideration: (a) measures developed for a single research project only; (b) measures designed for a narrowly specified population/occupational group (e.g., pharmacy students); (c) measures targeted toward younger age groups; (d) measures requiring special apparatus for administration; (e) measures requiring unusually long testing times; (f) measures requiring difficult or subjective scoring; and (g) measures requiring individual administration.

The result of this screening process was a second and more manageable list of candidate measures. Each measure on "List 2" was evaluated on 12 factors, listed in Figure 9, by at least two knowledgeable members of the research staff. (A five-point rating scale of potential usefulness was used to rate each of the 12 factors.) These ratings were used to guide the selection of the measures for the third list. However, this list ("List 3") still contained too many measures to administer in the time available. Therefore, List 3 was subjected to a final review by Project A researchers with the emphasis placed on "best bets" for prediction of on-the-job performance, given their collective knowledge of the constructs measured by the potential predictors and the factors that make up the criterion space.

The final content of the Preliminary Battery was a set of eight, timed, cognitive ability tests; a biographical questionnaire; a five-scale personal opinion inventory; and a vocational interests inventory. These instruments, collectively, measure a large number of human attributes not currently tapped by preinduction testing. The instruments in the Battery include:

1) ETS Figure Classification
2) ETS Map Planning
3) ETS Choosing a Path
4) ETS Following Directions
5) ETS Hidden Figures
6) ETS Space Visualization
7) EAS Numerical Reasoning
8) FIT Assembly
9) Owens' Biographical Questionnaire
10) Conscientiousness Scale
11) Stress Reaction
12) Leadership (Social Potency)
13) Motivation (Locus of Control)
14) AWOL/Delinquency (Socialization)
15) Vocational Occupational Interest Career Examination

1. Discriminability--extent to which the measure has sufficient score range and variance, i.e., does not suffer from ceiling and floor effects with respect to the applicant population.

2. Reliability--degree of reliability as measured by traditional psychometric methods such as test-retest, internal consistency, or parallel forms reliability.

3. Group Score Differences (Differential Impact)--extent to which there are mean and variance differences in scores across groups defined by age, sex, race, or ethnic groups; a high score indicates little or no mean differences across these groups.

4. Consistency/Robustness of Administration and Scoring--extent to which administration and scoring is standardized, ease of administration and scoring, consistency of administration and scoring across administrators and locations.

5. Generality--extent to which predictor measures a fairly general or broad ability or construct.

6. Criterion-Related Valdity--the level of correlation of the predictor with measures of job performance, training performance and turnover/attrition.

7. Construct Validity--the amount of evidence existing to support the predictor as a measure of a distinct construct (correlational research, experimental research, etc.).

8. Face Validity/Applicant Acceptance--extent to which the appearance and administration methods of the predictor enhance or detract from its plausibility or acceptability to laymen as an appropriate test for the Army.

9. Differential Validity--existence of significantly different criterion-related validity coefficients between groups of legal or societal concern (race, sex, age); a high score indicates little or no differences in validity for these groups.

10. Test Fairness--degree to which slopes, intercepts, and standard errors of estimate differ across groups of legal or societal concern (race, sex, age) when predictor scores are regressed on important criteria (job performance, turnover, training); a high score indicates fairness (little or no differences in slopes, intercepts, and standard errors of estimate).

11. Usefulness of Classification--extent to which the measure or predictor will be useful in classifying persons into different specialties.

12. Overall Usefulness for Predicting Army Criteria--extent to which predictor is likely to contribute to the overall or individual prediction of criteria important to the Army (e.g., AWOL, drug use, attrition, unsuitability, job performance, and training).

FIGURE 9. Factors Used to Evaluate Predictor Measures for the Preliminary Battery

The development of computerized selection measures in the perceptual and psychomotor domains is a special emphasis of this project. (Computer-adaptive testing, as that term is usually employed, is being amply pursued by other military research projects and is not our primary focus.) Accordingly, we conducted several activities to get an early start on this part of the project. First, we visited four military laboratories or field units where currently active research using such computerized measures was underway. Second, we developed a demonstration battery of computerized measures on a portable microprocessor (an Osborne 1) to become familiar with software and hardware problems. Finally, we reviewed the output of the literature search described above as regards the reliability and validity both of computerized measures for personnel selection and psychomotor/ perceptual tests.

The four site visits were the Air Force Human Resources Laboratory, Brooks Air Force Base, TX; the Naval Aerospace Medical Research Laboratory, Pensacola Naval Station, FL; the Army Research Institute Field Unit at Ft. Rucker, AL; and the Army Research Institute Field Unit at Ft. Knox, KY. During these visits we tried to answer five questions. The questions and the answers we obtained can be summarized as follows:

(1) What computerized measures are in use?

We found over 60 different measures in use across the four sites. A sizable number of these were evaluated via specialized simulators that are not relevant for this project (e.g., a helicopter simulator weighing several tons that is permanently mounted in an air-conditioned building). There were, however, many measures in the perceptual, cognitive, and psychomotor areas that were relevant.

(2) What computers were selected for use?

(3) What computer languages are being used?

We observed three different microprocessors in use: Apple, Terak, and PDP 11; and three different computer languages: PASCAL, BASIC, and FORTRAN. There appears to be relatively little in common among the four sites in terms of the hardware/software used.

(4) How reliable are these computerized measures?

(5) What criterion-related validity evidence exists so far for these measures?

Data are currently being collected at all four sites to address the reliability and criterion-related validity questions. The research at AFHRL is at the point of administering computerized measures to fairly large samples of subjects. This is also true of the research at Ft. Rucker where they expect to have validity data collected and analyzed later this year. Documentation of the results of these efforts will allow estimation of

the reliability and criterion-related validity of the measures under examination at these two locations.

A number of the measures have been under research at NAMRL for some time now, but criterion-related validity has not been the primary focus of that research. The prototype information processing measures developed there have been shown to be sensitive to individual differences within chronological age groups as well as to age-related changes across different age groups.

Data on the computerized measures at Ft. Knox are currently being analyzed. While there are potential problems with range restriction in the predictors and the criterion measures, they are finding significant, positive correlations between micro-processor measures and their higher fidelity, "hands-on" counterparts.

To summarize, there is little information currently available regarding the reliability or criterion-related validity of the computerized measures in use at these sites. This is not surprising because most of these measures have been developed only recently.

After conducting these site visits, we programmed a short demonstration battery in BASIC on the Osborne 1, a portable microprocessor. The purpose of this activity was to implement some of the techniques and procedures observed during the site visits to determine the degree of difficulty of such programming and to get an early gauge on the quality of results to be expected from use of a common portable microprocessor and a language that is common to many machines, but which has some disadvantages in terms of processing power, speed, and flexibility. This short battery was self-administered, recorded the response and time to respond, and contained five tests: simple reaction time (pressing a key when a stimulus appeared), choice reaction time (pressing one of two keys in response to one of two stimuli), perceptual speed and accuracy (comparing two alphanumeric phrases for similarity), verbal comprehension (vocabulary knowedge), and a self-rating form (indicating which of two adjectives "best" describes the test taker, on a relative seven-point scale). We also experimented with the programming of several types of visual tracking tests, but did not include these in the self-administered demonstration battery.

Summary of First Year Activities

In sum, we have accomplished what we think is a landmark survey of potential selection measures for improving selection and classification decisions for U.S. Army enlisted personnel. Based on this survey, additional analyses of expert judgment, and several reviews, we developed the preliminary selection battery, which has been carefully designed to provide comprehensive information about what kinds of measures will provide the most useful supplements to the ASVAB. Finally, we have begun the initial development work for new psychomotor and perceptual tests that could become part of the preinduction test battery.

Next Steps in the Development of Selection Predictors

During the next year the following activities will be carried out for the purpose of developing and validating new selection measures.

1. The Preliminary Battery will be administered to trainees entering classes in the following MOS: 05C (Ft. Gordon), 19E/K (Ft. Knox), 63B (Ft. Dix and Ft. Leonard Wood), and 71L (Ft. Jackson). Testing monitors and administrators at each site were trained during September 1983, and testing will be carried out from October 1983 through June 1984.

2. A technical review of possible predictor measures will occur in October. This will consist of collecting and analyzing expert judgments of the expected relationship between the most promising predictor constructs and the various performance factors in the training, Army wide, and MOS-specific performance domains.

3. Development of computerized measures will continue, including a pretest in November and January at a MEPS station. A preliminary report on computerized measures will be prepared in March.

4. In March, the measures to be included in further development will be selected (based primarily upon the technical and cost reviews mentioned in points 2 and 3 above). These measures will then be known as the Pilot Trial Battery. Item writing will begin and tryouts are scheduled for March, April, and May of 1984.

5. Initial data from the Preliminary Battery administration will be analyzed in January and February, and the results will be used to inform Pilot Trial Battery development (described in point 4).

6. In June 1984, the Pilot Trial Battery will be put into final form for the pilot test in summer/fall of 1984 in CONUS.

Abstracts

As noted in the Introduction, abstracts of relevant and related research reports follow:

ASSESSMENT OF PRACTICE EFFECTS: TEST-RETEST SCORES FOR FY81 ACTIVE ARMY APPLICANTS ON ASVAB 8/9/10*
D. Friedman, A. Streicher, H. Wing and F. Grafton
(ARI)

Approximately 30,000 FY81 active Army applicants with initial and retest scores on ASVAB 8/9/10 were identified. For each of the 10 subtests of the ASVAB, score changes were investigated by sex of applicant. Alternate forms reliability for each subtest was high. Factor analyses of the subtest scores replicated analyses performed with initial scores of other Army applicants.

EXAMINATION OF ABILITY REQUIREMENTS FOR THE INFANTRY CAREER MANAGEMENT FIELD**
D. M. Olson and L. M. Hanser
(ARI)

This exploratory research examined whether specific cognitive, perceptual, or psychomotor abilities could be identified as requirements for effective performance in the Infantry Career Management Field (CMF 11). A new computerized Job Assessment System was used to rate the ability requirements for the Infantry MOS.

Requirements for the Infantry MOS were scaled by six company commanders and three NCO, who were stationed in training units at Ft. Benning, GA. These male raters provided a total of 21 sets of ratings. The computerized rating format presented conceptual definitions for 39 general abilities via visual display on a CRT. Once the ability was judged as required for performance, raters assigned a value on a seven-point scale, which indicated the magnitude of the ability required. Profiles of ratings for each Infantry MOS were obtained by averaging the ratings of all SME who evaluated the same MOS.

Comparisons of ability requirements across Infantry MOS indicated that Memorization and Spatial Orientation were rated as highly required for successful performance in the Infantry, while such abilities as Static and Trunk Strength and Rate Control received consistently low ratings. Application of this computerized methodology could describe the ability requirements of various military jobs, and establish linkages between ability taxonomies and duties of specific MOS.

*Paper presented at the 24th Annual Conference of the Military Testing Association in San Antonio, Texas, November 1982.
**Paper presented at the 25th Annual Conference of the Military Testing Association in Gulf Shores, Alabama, October 1983.

VERBAL INFORMATION PROCESSING PARADIGMS: A REVIEW OF THEORY AND METHODS*

K. J. Mitchell
(ARI)

The theory and methods of selected verbal information processing paradigms were reviewed. Work in factor analytic, information processing, chronometric analyses, componential analyses, and cognitive correlates psychology were discussed.

The definition and measurement of cognitive processing operation, stores, and strategies involved in performance on verbal test items and test-like tasks were documented. Portions of reviewed verbal processing paradigms were synthesized and a general model of text processing was presented.

The verbal processing model served as a conceptual framework for the subsequent identification and assessment of cognitive processing contributions to performance on the verbal subtests of ASVAB 8/9/10. These results were also used in a series of analyses on the predictive validity of assessed constructs to successful performance in Army training.

FINAL STATUS REPORT ON THE COMPARABILITY OF ASVAB 6/7 AND 8/9/10 APTITUDE AREA SCORE SCALES**

F. C. Grafton, K. J. Mitchell, and H. Wing
(ARI)

Preliminary analyses of Defense Manpower Data Center (DMDC) and Military Enlistment Processing Command (MEPCOM) accession data for FY78-81 point to discrepancies in the distributions of aptitude area scores for ASVAB 7/6 and ASVAB 8/9/10. Of concern are: (1) the equivalence of Armed Forces Qualification Test (AFQT) and aptitude area scores assigned through the recalibrated ASVAB 8/9/10 testing program, and (2) the quality of applicants accessed during Fiscal Years 1977-1982. These issues were addressed, in part, in the Interim Status Report on the Comparability of ASVAB 6/7 and ASVAB 8/9/10 Aptitude Area Score Scales, by Grafton, Mitchell, and Wing (ARI Selection and Classification Technical Area Working Paper 82-5). This final report addresses these concerns more completely.

The data show an increase in accession quality on AFQT and the aptitude area composites during Fiscal Years 1977-1982. The FY82 accessions had mean AFQT and aptitude area scores above the population means; this difference reflects the rise in accession standards instituted in March 1982. FY81 accessions taking ASVAB 8/9/10 had slightly lower AFQT and aptitude area score means than those in FY82. The means and cumulative frequency plots for FY81 ASVAB 6/7 accessions, with potential ineligibles excluded, were comparable to the ASVAB 8/9/10 data for AFQT and nine of the aptitude areas. Mean scores for CY76-78 and FY81 ASVAB 6/7 accessions with records of potential ineligibles included in the samples were below the population means.

*To be published as an ARI Technical Report.
**To be published as Selection and Classification Technical Area Working Paper 82-6.

XI. SUMMARY AND CONCLUSIONS

Project A was designed and is being executed within a different framework than that of previous R&D projects in the behavioral and social sciences. Historically, past practice has been to allocate funds in relatively small amounts to one investigator or to one research firm for a relatively circumscribed piece of work. Project A was conceived differently. Its aim is to address an integrated set of R&D questions and problems within one project, to develop a complete personnel system for selecting and classifying all entry-level personnel in a large organization. Much of the information required to develop such a system could not be produced by a set of piecemeal projects. Consequently, while the magnitude of Project A is large in terms of total funding, time frame, size of the research staff, and number of research participants, it is expected to produce much more information in a shorter time than would have been the case if the usual framework for allocating R&D funds had been allowed.

Executing and managing such a large integrated project place heavy responsibilities both on the contractor staff and on the professional staff of the Army Research Institute. However, concomitant with the burden of responsibility is the expectation that the resulting classification procedure will be grounded in the most complete data base ever developed for a large personnel system and that many of the most vexing research questions in the field will be addressed comprehensively and directly. After one year's experience on the project, the weight of responsibility and the realization that the payoffs will far exceed anything that has gone before, are felt more intensely than ever.

Planning Activities

In general, the first year's activities have been taken up by an intensive period of detailed planning, briefing the advisory groups, preparing the initial troop requests, and beginning the comprehensive predictor and criterion development that will be the basis for the later validation work. The requirement for a detailed research plan to be produced during the first six months of the contract was included in the RFP; hindsight judges it to be an even more valuable step than the authors of the RFP might have had in mind. The research staff devoted a great deal of effort to the writing of the research plan, it was carefully reviewed by the advisory groups and by the ARI professional staff, revisions were made, and the completed plan was published in May 1983 under the joint authorship of the contractor and ARI staffs. The Research Plan and the accompanying Master Plan carefully lay out, in detailed fashion, the specific steps to be taken by each subtask in the project, the schedule which will be followed, and the budget allocations that will be made to each subtask during each contract period. These two documents have become the guiding blueprint for the project. They have also proven invaluable as a mechanism for developing a consensus and facilitating communication among contractor staff and between the contractor and ARI.

The detailed planning and review that went into the development of the Research Plan and Master Plan made it possible to lay out clearly and

precisely the troop support the project would need during its first two years. Consequently, the project has experienced relatively little difficulty in communicating its needs to the appropriate Army organizations and in gaining their support. For the outstanding cooperation we have received so far, we are most appreciative.

Substantive Activities

The previous chapters in this report have outlined and briefly discussed the substantive activities that have taken place during the first contract year. The major points from this discussion are summarized below.

The LRDB and the FY81/82 Data File

As noted by Hanser and Grafton (1983), no one should expect easy going when attempting to use large-scale computer files of archival data for personnel research. Computerized information systems in large organizations are designed to serve purposes other than personnel research. Consequently, it came as no surprise that the predictor and criterion data for FY81/82 accessions were not as neat and clean as we would have liked. In fact, a tremendous amount of effort was devoted to obtaining and merging computer files, editing records, and filling gaps in documentation. However, the result has been the creation of the most extensive file of archival records that has ever been generated for purposes of personnel research. The files encompass two years of Army accessions (approximately 200,000 people drawn from approximately 500,000 applicants and subsequently placed in over 300 different skilled entry-level positions). While the available edited records fall short of containing complete data for everyone, the magnitude of the data base is considerable.

Work on the analyses of these data has just begun and it is too early to make definitive statements about empirical findings. However, one principal objective is to use 81/82 data files to investigate the validity of new or revised composites of ASVAB subtests to predict success in a wide variety of job training schools. Regardless of whether the analyses point toward new composites, revisions, or no change in those currently being used, the analyses will be based on a far larger data base than ever before.

One obvious, but extremely important, finding from preliminary analyses is that although the Army is a large organization it is not so large that every MOS (job specialty) contains a sufficient number of incumbents to permit statistical validation analyses. It simply will not be possible to estimate prediction equations empirically for every MOS. Also any validation analysis must deal with differing criterion metrics across MOS, restriction of range due to selection, and considerable skewness in the criterion distributions. Consequently, one extremely significant outcome resulting from having the FY81/82 data file is that alternative analytic metrods for dealing with these problems can be tried out and evaluated, so that the analyses of the FY83/84 and FY86/87 data can proceed efficiently and appropriately. The analyses of the FY81/82 data base will serve as the benchmark with which the subsequent results to be produced by Project A can be compared. That

is, we now have enough information in the FY81/82 file to provide a reasonably clear picture of how much selection validity and classification efficiency can be produced within the current system using the current data base.

MOS Task Descriptions

Because the Army's MOS job analysis information was not generated for personnel research purposes, the analysis data needed considerable modification before they could be used by Project A for criterion development. Consequently, a great deal of effort was devoted to refining and integrating task descriptions from the Soldier's Manual and the CODAP occupational survey questionnaires.

For each MOS, a data bank of task statements was accumulated from all available sources, and the individual task statements were edited to determine if they indeed focused on observable job tasks, if they were redundant or overlapped with other tasks, and if they were at the same level of generality. Subject matter experts were used to determine if the edited pool of task descriptions provided a complete picture of the content of the MOS. The SME also judged the relative criticality of each task. These steps are currently being carried out for focal MOS so that there will be a precise and thoroughly developed task description for the MOS being considered in Project A. The task descriptions will provide the principal basis for the development of hands-on performance measures and job knowledge tests. As such, they should provide a much better foundation for the subsequent criterion development than has been available in the past.

Assessment of Training Performance

A major objective of Project A is to use a comprehensive, valid, standardized test construction procedure to develop a measure of training success for each focal MOS so that the item content represents both the content of training and the content of the job. That is, the items will sample the job content representatively and will be further identified as being covered in training versus not being covered in training. When this is accomplished, a measure of direct learning in training (items that match training content) can be related to a variety of job performance criteria with and without ability (as measured by predictor tests) controlled.

To meet these objectives the project staff have spent the last several months visiting key training schools and developing job task descriptions for each MOS. What has been produced is a thorough analysis of the objectives, curriculum, and assessment procedures for the key schools. The process of describing the job content and matching it with training content has just begun and will be completed during FY84. When the matching of training content and job content is completed and the knowledge tests are constructed, we will have achieved the capability for determining how training performance is, or is not, related to job performance.

Job Performance Criterion Development

Our initial model of soldier effectiveness was preliminary, saying essentially that both specific task performance and the general factors of commitment, morale, and organizational socialization comprised the total domain.

During this year the task descriptions for the four MOS in Batch A have been completed and Batch B is in progress. Further, virtually all the critical incident workshops for MOS-specific task performance factors have been completed. This easily has been the most massive effort ever undertaken to apply these methods to criterion development. There now exist hundreds of critical incidents of specific task performance within each focal MOS, and thousands of critical incidents describing performance behaviors that have a general, not MOS-specific, referent. These large samples of job behaviors are being used to identify MOS-specific and MOS-general performance factors and to develop rating scales (during FY84) to assess individual performance on these factors. This process has produced a revised and expanded model of the criterion space that will be used to generate further criterion development work and to guide predictor selection.

An additional important outcome of the interaction between model development and task/behavior description is the identification of an array of MOS-specific task performance factors that are intended to encompass the unique task content of all MOS in the enlisted personnel job structure. Although it is only a first cut, it will be the basis for the further development of a standardized set of task descriptors that can be applied to any MOS so as to describe thoroughly its content. Such a standardized measure will make it possible to answer a number of important questions that could not have been addressed previously. For example, how similar precisely are any two MOS in terms of their job content? Should they have a common selection algorithm? How different should their training schools be?

Predictor Selection

A major objective that had to be accomplished during the first contract year was to select the preliminary predictor battery for administration to the 83/84 longitudinal sample and to lay the groundwork for the development of the trial predictor battery. To do this, the project staff carried out what was perhaps the most massive literature search ever done in personnel psychology. The result has been: (a) a very thorough and precise description of the specific measures that might be useful in any selection or classification effort; (b) a summary of the empirical evidence attendant to each one; and (c) an explication of the latent variables, or constructs, that seem best to represent the content of the operational measures or tests.

The value of this information, while it is extremely high for this particular project, goes far beyond the boundaries of Project A. It will be of crucial importance for almost any personnel selection project that comes after, regardless of the specific jobs or organizations in question. There is now a wealth of valuable and well-organized information that is available for use in future work.

In Conclusion

During its first year of full scale operation with ARI and consortium scientists, Project A has stayed on schedule and within its budget. More attention was devoted to detailed planning and outside review than originally envisioned. However, these very thorough and careful preparatory steps were well worthwhile in terms of facilitating communication among everyone associated with the project and uncovering all the unresolved issues that would have plagued us at some later time. Most importantly, it served to coalesce all of the diverse organizational elements whose informed cooperation was essential to the successful execution of the research program. Project A has indeed become a unified and integrated effort.

Also, although much of the research activity during the first year was designed as essentially preparatory, some valuable first year products include the 81/82 data file, the task banks, the critical incident banks, and the literature review of the predictor domain.

We look forward to a productive second year.

REFERENCES

Bork, R. D. & Mislevy, R. J. (1981). Data quality analysis of the Armed Services Vocational Aptitude Battery. Chicago: National Opinion Research Center.

Eaton, N. K. & Goer, M. H. (1983). Improving the Selection, Classification, and Utilization of Army Enlisted Personnel: Technical Appendix to the Annual Report. ARI Research Note 83-37. Alexandria, VA.

Goer, M. H., Sadacca, R., & Harris, J. H. (1983). Development and Validation of Army Selection and Classification Measures. Project A: Integrated Master Plan. Alexandria, VA: Human Resources Research Organization.

Hanser, L. & Grafton, F. (1983). Dusting off old data: problems encountered with archival records. Paper presented at the American Psychological Association, Anaheim, CA.

Human Resources Research Organization, American Institutes for Research, Personnel Decisions Research Institute, and Army Research Institute. (1983). Improving the Selection, Classification, and Utilization of Army Enlisted Personnel. Project A: Research Plan. ARI Research Report 1332. Alexandria, VA.

James, L. R., Muliak, S. A., & Brett, J. M. (1982). Causal analysis: Assumptions, models, and data. Beverly Hills, CA.

James, L. R., Wolf, G., & Demaree, R. (1981). Estimating interrater reliability in incomplete designs (IBR-81-14). Arlington, VA: Organizational Effectiveness Research Programs.

Motowidlo, S. J., Dowell, B. E., Hopp, M. S., Borman, W. C., Johnson, P. D., & Dunnette, M. D. (1976). Motivation, satisfaction, and morale in Army careers: A review of theory and measurement (ARI Tech. Rep. TR-76-A7). Arlington, VA: U.S. Army Research Institute for the Behavioral and Social Sciences, (NTIS No. AD-A036390).

Porter, L. W., Steers, R. M., Mowdray, R. T., & Boulain, P. V. (1974). Organizational commitment, job satisfation, and turnover among psychiatric technicians. Journal of Applied Psychology, 59, 603-609.

Shrout, P. & Fleiss, J. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86, 420-428.

Smith, P. C. & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 47, 149-155.

Steers, R. M. (1977). Antecedents and outcomes of organizational commitment. Administrative Science Quarterly, 22, 46-56.

Van Mannem & Schein, E. H. (1979). Toward a theory of organizational socialization. In B. M. Shaw (ed.). Research in organizational behaviors (volume 1). Greenwich, CT: JAI Press.